The Thesis Committee for Sydney Kase
Certifies that this is the approved version of the following Thesis:

Texas Groundwater Conservation District Policy: Content Mining and Statistical Analysis

APPROVED BY
SUPERVISING COMMITTEE:

Sheila M. Olmstead, Supervisor

Charles W. Kreitler

Varun Rai

Texas Groundwater Conservation District Policy:
Content Mining and Statistical Analysis

by

Sydney Kase

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2019

Dedication

This thesis is dedicated to my family, Jeff, Liz, Niko, Rowan, and Ryan. Thank you for your love, support, and wisdom. Without you this would not have been possible.

Acknowledgements

Thank you to Dr. Olmstead, Dr. Kreitler and Dr. Rai for their support and guidance. I would like to thank the Texas Alliance of Groundwater Conservation Districts for their assistance and data. Thank you to the TWDB and TCEQ for their fantastic resources. Thank you to the Jackson School and the Energy and Earth Resources Program for making this possible.

Abstract

Texas Groundwater Conservation District Policy: Content Mining and Statistical Analysis

Sydney Kase, MA

The University of Texas at Austin, 2019

Supervisor: Sheila M. Olmstead

Groundwater is an increasingly significant and precious commodity within the state of Texas. The only statewide regulatory vehicle for governance and management of the groundwater resources are the Groundwater Conservation Districts (GCDs). A comprehensive statewide planning process was established by two senate bills in 1997 and 2001 which set forth the required actions for districts to manage and conserve the groundwater resources within the State of Texas. The bills require that all water conservation districts (including groundwater conservation districts, underwater conservation districts and subsidence districts) develop a management plan and update it at regular intervals. The management plans include a full accounting of the district's water demands and the water supplies, the resultant water need (shortage or surplus) within each district as well as the rules of the district. Each district's management plans are also required to establish a set of goals that the district will use to manage its water resources in order to meet its reported shortage or maintain a surplus water budget. GCDs are mandated to produce management plans during their initiation, as well as periodic updates over time.

In order to understand if the current management plan structure is working, I used content mining to turn the management plans into a dataset and then ran a series of statistical models to describe impacts. This research outlines a method of quantitative analysis to understand the relationship between groundwater management plans and groundwater resources that utilizes current and historic GCD management plans, and a water supply need metric developed by the Texas Water Development Board (TWDB). Statistical classification techniques were employed to evaluate the association between the management plans and the water supply class of each GCD. The statistical learning methods returned between 75% and 90% correct classifications depending on the model. The most impactful predictors when determining class were found to be shortage, recharge and groundwater when classifying as a surplus and precipitation, demand and aquifer when classifying a shortage.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1: PROJECT BACKGROUND

1.1 Research Approach, Background and Purpose

In a basic sense, policymakers and scientists tend to take different approaches when it comes to resource management. Policy is typically forward-facing, the term ‘management plan’ refers to a document that plans for the future. Science, especially geoscience, tends to look backward for patterns that help to understand the current environment. The objective of this research is to marry basic components of these two fields in order to develop an approach to resource management that looks over historical patterns in policy in order to understand its impact. To do this I applied quantitative analysis approaches to examine the interactions between policy documentation and resource supply status. The goals of this project are as follows:

1. Determine whether Groundwater Conservation District (GCD) management plans can be used to classify GCDs by water supply need in order to understand the association between management documents and water supply need.
2. Discover what components have the strongest impact on classification.

In order to achieve the two goals, I transformed policy documents into data tables through the use of data mining and then constructed a series of statistical models to test the relative impacts of policy and to find the best indicators of that impact.

The subject of this project is the management policy of groundwater conservation districts (GCDs). These policies were chosen because of the importance of groundwater as a resource, and the difficulty in evaluating the current policy landscape. Evaluating historical GCD policy is difficult for several reasons: the relatively new nature of GCD management plans, the lack of standardization across districts, the historical inaccessibility of policy documentation and the absence of qualitative metrics for assessing management plan impacts. The oldest GCD management plan in the state of Texas is only about two decades old, which means that the implemented policies really haven't had much time to have an impact. To avoid the pitfalls of judging a GCD's resource by a policy that has not had enough time to bear fruit, this project uses a cumulative annual water supply metric to tie the time-period of the policy directly to the net change in water supply over the course of the time that the policy has been effective. Texas' groundwater districts are very diverse not only in terms of physical formation, but also infrastructure, user group composition, withdrawal rates, recharge rates, water level measurement coverage and frequency, well counts and permitting, and of course management and policy. Fortunately, the TWDB's water supply need metric looks at every district as equal, so this research was able to leverage that standardization to establish a basic, common metric for every GCD. Lastly, the inaccessibility of GCD policy documentation was circumvented by using data-mining techniques to transform policy management document content into tabular data. Through this process I hope to assist current and future policymakers at the statewide level by providing insight into the relationship between the GCD management structure and the water supply needs of the GCDs.

1.2 Data Mining: What is it?

Data Mining is a relatively new phenomenon that has taken hold in both the academic world and the corporate world. It is the process of examining text or image content and extracting information in order to categorize useful information or discover patterns. The practical applications for this technique are vast, but some notable examples include assisting with business decisions, understanding how networks form and function, and developing data and information requirements. The major tenets of Data Mining can be broken down into the following steps: First, extract, transform and load data in a functional format, Next, assemble the converted content into a database. The third step is data analysis, which is very flexible depending on the use case and end goals. Key tenants of data mining analysis include pattern predictions using trends, outcome predictions based on recorded behavior and clustering based on findings. The last step is to present the results of the analysis in a concise, understandable form so that they can be used to gain insight into the subject content.

1.3 The State of Groundwater in Texas

1.3.1 The Rule of Capture (Legal framework for groundwater management)

For over one hundred years Texas has used a law called the Rule of Capture to determine groundwater rights. Unlike surface water, which mainly belongs to the state unless it falls on federal land, groundwater in Texas belongs to the owner of the land directly above it. The case that established the rule of capture in Texas was Houston & Texas Central Railroad Co. v. East in 1904 (Texas Supreme Court, 1904). It involved a landowner (East) who had a well run dry due to the extraction of several thousand gallons of water per day by an adjacent railroad company (Houston & Texas Central Railroad Company). The landowner sued the railroad company in an attempt to recover damages. In that case, the court decided to use the rule of capture law instead of the law of “reasonable use.” Reasonable use would stipulate that landowners do not have an absolute right to groundwater, but rather that withdrawal is limited to only the necessary amount required for reasonable use of the land (Texas Supreme Court, 1978). Today, Texas is the only remaining U.S. state that follows the rule of capture for groundwater. Unfortunately, when the rule was established there was limited scientific understanding of the way that water behaves under the ground. The wording of the court's decision in the 'East' case referred to groundwater as "secret, occult and concealed", and asserts that the administration of the law concerning surface water rights could not be applied to groundwater (Mullican, 2004). Fortunately, since the 1900s the law has been updated to outlaw malevolent damage, deliberate disposal, and effective subsidence via groundwater withdrawal. Further, the Texas Legislature passed the Conservation Amendment in 1917 (Texas Const. art. XVI, §59), which grants the ability to regulate groundwater to the Legislature. The Texas Supreme Court continues to reiterate the legislature's ability to control groundwater regulation, but the Legislature has yet to exercise that directive statewide. If this ever were to happen, it would fundamentally make the rule of capture obsolete. Rule of Capture law still allows an individual landowner or company to build wells at increasing depths, creating a scenario where the deepest well can draw down the level of the water table so much that it causes all surrounding wells to go dry without any legal ramifications. Because Texas courts do not directly regulate groundwater from overuse due to the rule of capture, groundwater must be managed by other means. Groundwater Conservation Districts are those other means; they are governing bodies that exist to regulate groundwater withdrawal and manage groundwater resources and infrastructure.

1.3.2 WHERE IS THE WATER AND WHY DO WE CARE?

Groundwater is a significant and important commodity within the state of Texas and is expected to continue to be a water resource that meets projected water supply needs growth. The 2017 TWDB Texas State Water Plan reports that some form of groundwater use will provide 11.29% of Texas water supply by volume in 2020 (Bruun, 2017). According to a separate TWDB technical report released in 2015 (Neffendorf, Hopkins, 2015), groundwater levels dropped throughout the state between 2013 and 2014, in keeping with the overall historical trend. A higher percentage of groundwater wells showed declines in water level at the end of 2014 than at the end of 2013. The median water level change between 2013 and 2014 for the major aquifers was a decline of 1.2 feet with an accompanying 73 percent of wells showing declines in water level. In contrast, 68 percent of wells saw a water level decline between 2012 and 2013, 75 percent of wells saw a water level decline between 2011 and 2012 and a spike of 92 percent of wells saw a water level decline between 2010 and 2011 (Neffendorf, Hopkins, 2015).

Table 1: Summary of median water-level changes by major aquifer and region.

Major Aquifer	Region	Number of Wells	Median Change (feet) 2013-2014	Median Change (feet) 2012-2013	Median Change (feet) 2011-2012	Median Change (feet) 2010-2011
Trinity	Central	41	-2.4	-0.1	-0.9	-16.7
Hueco(-Mesilla) Bolsons	West	1	-2.2	-0.4	-3.5	1.5
Pecos Valley	West	4	-1.7	-0.9	-0.6	-7.6
Edwards (Balcones Fault Zone)	Northern Central	4	-1.4	12.1	0.9	-3.5
Ogallala	High Plains	26[a]	-1.2	-1	-1.8	-1.9
Seymour	Rolling Plains	2	-1	-0.8	-0.9	-3.2
Trinity	Northern Central	16	-0.9	-0.8	-1.6	-8.5
Edwards-Trinity (Plateau)	West	24[b]	0.5	-0.8	-0.9	-0.7
Gulf Coast	South and East	13[c]	0.7	-0.9	0.5	-6.3
Carrizo-Wilcox	South and East	12	1.5	-0.2	-0.9	-4.4

a. Three wells in Dallam, Armstrong, and Terry counties replaced three wells taken out of service in Armstrong, Carson, and Dawson counties

b. Three wells added in Schleicher County

c. Two wells added in San Jacinto and Polk counties

Source: Summary of Groundwater Conditions in Texas: Recent (2013–2014) and Historical Water Level Changes in the TWDB Recorder Network, By Blake Neffendorf and Janie Hopkins. December 2015

The spikes in median water level decline and percentage of declining water level wells between 2010 and 2011 are likely related to the significant drought that occurred over that time period. No similar environmental factor explains the higher water level declines recorded between 2013 and 2014. As recorded in Table 1, the statewide median water-level change from 2013 to 2014 in all major aquifers was -1.2 feet. Overall the state's aquifers are experiencing yearly water level declines and the year-to-year change in the areas of the state that are recording water level rises is less than the annual difference for areas that are experiencing decline.

Groundwater already accounts for a significant amount of the more than 16 million acre-feet of water used annually in the state of Texas. The reliance on groundwater and future projected expansion is due to several factors. Current surface water supplies are not amenable to growth without substantial infrastructure investment in the form of new reservoirs, dams, etc., and land acquisition is a significant factor. Several of the non-groundwater strategies involve some reliance on conservation measures, which are limited to a much smaller total volume with current available technologies.

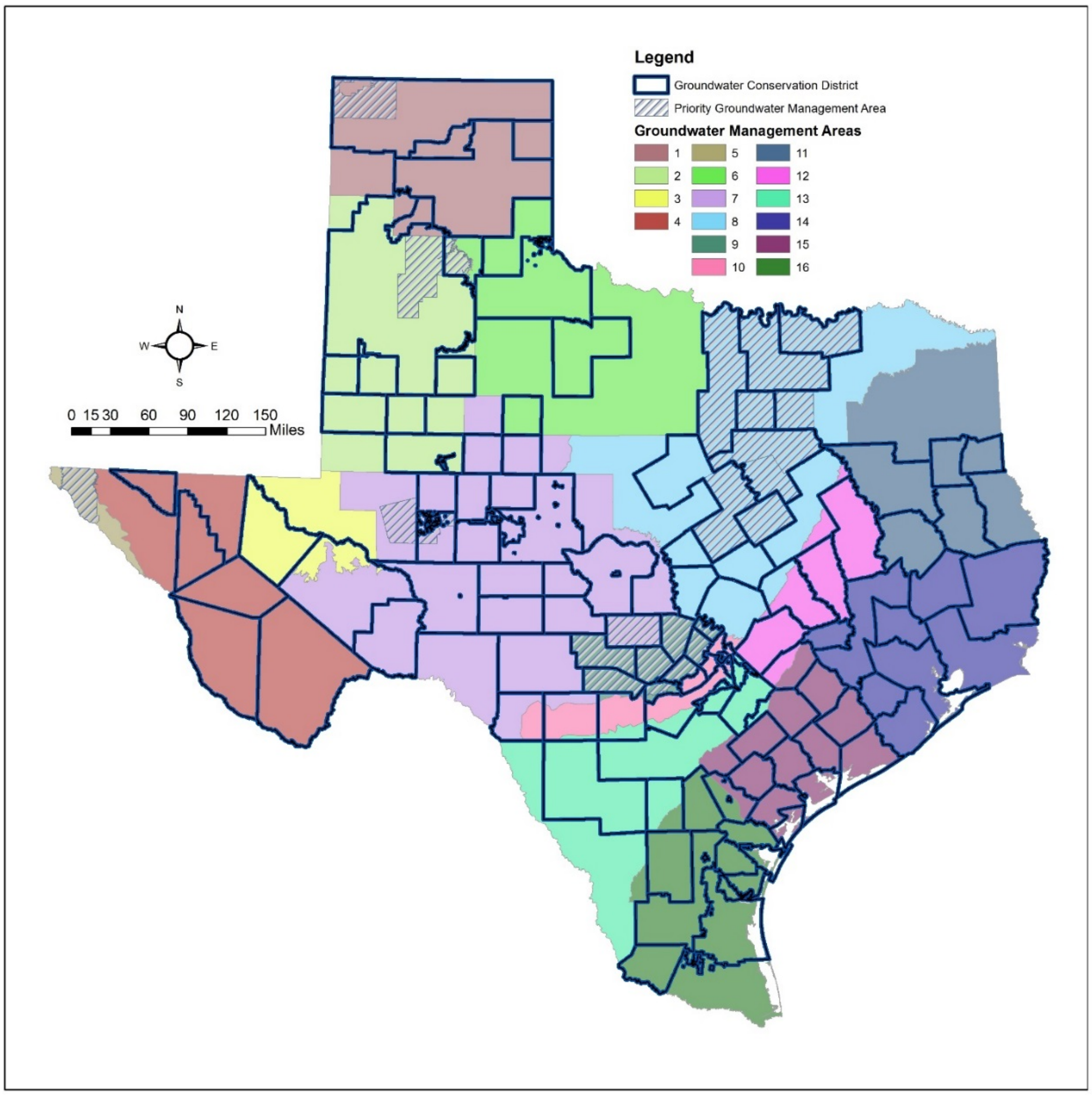
1.3.3 HOW IS GROUNDWATER MANAGED?

The Texas Legislature created Groundwater Conservation Districts to protect and maintain the groundwater of Texas. The districts collectively serve 16 Groundwater Management Areas (GMAs), which were established in 1995 by the 74th Texas Legislature[1]. The TWDB was granted the responsibly of delineating the boundaries of the GMAs and elected to use the boundaries of the major aquifers of Texas as a guide. In areas of the state where multiple major aquifers coincided, the TWDB largely deferred to the boundary of the shallowest aquifer for GMA delineation. Because of this deference, a handful of major aquifers are split into multiple GMAs. Aquifer divisions were decided in accordance with natural features, as well as groundwater withdrawal patterns and variations in hydrogeologic properties. Additionally, where possible, the TWDB aligned GMA boundaries with those of counties and existing GCDs. The primary purpose for the

[1] Section 35.004, Chapter 35, Title 2, Texas Water Code

delineation of GMAs is to facilitate joint planning by GCDs that manage the same aquifer.

Figure 1: A map of the Groundwater Conservation Districts of Texas, overlying the Groundwater Management Areas and Priority Groundwater Management Areas



Sources: Texas Water Development Board, Texas Commission on Environmental Quality

Seven Priority Groundwater Management Areas (PGMAs) have been established, covering all or part of 35 counties. According to the TCEQ;

> A Priority Groundwater Management Area (PGMA) is an area designated and delineated by TCEQ that is experiencing, or is expected to experience, within 50 years, critical groundwater problems including shortages of surface water or groundwater, land subsidence resulting from groundwater withdrawal, or contamination of groundwater supplies. (TCEQ, 2016)

Unlike GMAs, most PGMAs are delineated by county boundaries and are established by TCEQ. TCEQ can create special Groundwater Conservation Districts, but only within PGMAs. PGMAs can also be established by the Texas Parks and Wildlife Department, or the Texas Water Development Board. However, this action can only be taken for areas where it is established that there are "critical groundwater problems". The areas must be identified and studied first, and critical issues proved before a PGMA is established. PGMAs are designed to encourage the creation of GCDs and can mandate their creation if necessary.

The only statewide regulatory vehicle for governance and management of groundwater resources are the Groundwater Conservation Districts (GCDs). The comprehensive statewide planning process was established by two Senate bills in 1997 and 2001 which set forth the required actions for districts to manage and conserve Texas groundwater. The bills require that all water conservation districts (including groundwater conservation districts, underwater conservation districts and subsidence districts) develop a management plan and update it at regular intervals. A groundwater management plan establishes a GCD's groundwater management goals. The first groundwater management plan approved by the TWDB was the Gonzales County UWCD's management plan in 1998. Every GCD management plan has statutorily required elements, which are broken into two main categories; goals and information.

Table 2 below includes the required goals and information for all GCD management plans.

Table 2: Required Elements of a GCD Management Plan, TWDB 2017

Required Goals:	To provide the most efficient use of groundwater;
	To control and prevent waste of groundwater;
	To control and prevent subsidence;
	To address conjunctive surface water management issues;
	To address natural resource issues that impact the use and availability of groundwater, and which are impacted by the use of groundwater;
	To address drought conditions;
	To address conservation, recharge enhancement, rainwater harvesting, precipitation enhancement, and brush control, where appropriate and cost-effective; and
	To address the desired future conditions established pursuant to the Texas Water Code.
Required Information:	Performance standards and management objectives under which the GCD will operate to achieve its management goals;
	Details of how the GCD will manage groundwater supplies in the district, including a methodology by which the GCD will track its progress in achieving its management goals;
	Detailed descriptions of actions, procedures, performance and avoidance that are, or may be necessary, to affect the plan including specifications and proposed rules;
	Estimates of the following:
	• modeled available groundwater (MAG) in the GCD based on the desired future condition (DFC) established under Texas Water Code §36.108;
	• the amount of groundwater being used within the GCD on an annual basis;
	• the annual amount of recharge from precipitation, if any, to the groundwater resources within the GCD;
	• the annual volume of water that discharges from each aquifer in the GCD to springs and surface water bodies;
	• the annual volumes of flow into and out of the GCD within each aquifer and between aquifers in the GCD if a groundwater availability model is present;
	• the projected surface water supply in the GCD according to the most recent state water plan;
	• the projected total demand for water within the GCD according to the most recent state water plan; and
	Consideration of the water supply needs and water management strategies within the county(s) covered by the GCD according to the most recent state water plan.

Source: Texas Water Development Board

As is evident from Table 2, most of the GCD-mandated goals cast a wide net in order to allow each district to tailor their goals to the unique needs of their district. Some goals are expressed in very broad terms and others are outlined with very specific success measures. Since every district has its own distinct set of challenges, each district may have as few as one and as many as twenty management goals, objectives and performance standards. In 2005 the 79th Texas Legislature enacted HB 1763 that requires planning coordination among districts that are in the same GMA. Through this coordinated planning effort, districts must establish the Desired Future Conditions (DFCs) of the aquifers within their shared GMAs. DFCs are defined as;

> the desired, quantified condition of groundwater resources (such as water levels, spring flows, or volumes) within a management area at one or more specified future times as defined by participating groundwater conservation districts within a groundwater management area as part of the joint planning process. (Title 31, Part 10, §356.10 (6) of the Texas Administrative Code)

Desired future conditions are required to be physically possible on the individual GCD level, as well as attainable across the entire GMA. If there are discrepancies in the DFCs produced for different geographic areas overlying the same aquifer or subdivision of an aquifer, then the GCDs and other interested parties must work together to resolve the differences before the DFCs can be approved by the TWDB. The specified time period to achieve the DFC extends through at least the period that includes the current planning period for the development of regional water plans pursuant to §16.053, Texas Water Code, or in perpetuity, as defined by participating GCDs within a GMA as part of the joint planning process.

Groundwater Availability Models (GAMs) are the mechanism through which GCDs evaluate sustainable water levels for existing and future conditions. Each model is calibrated to ensure that it can reasonably reproduce past water levels and groundwater flows. GAMs are typically developed and maintained by consultant geologists and engineers and are preserved by the TWDB. They use current and historical water level data, geological information about the aquifer, and recharge data from precipitation and streamflow to estimate the characteristics of groundwater in the aquifer and predict the aquifer's water level under normal and drought conditions. There are currently 16 GAMs developed for Texas' nine major aquifers.

1.3.4 The Origins and Construction of Groundwater Management Plans

All GCDs are required to generate a groundwater management plan and submit it to the Texas Water Development Board (TWDB) for approval. When a new district is created, it must submit a management plan within three years. If the new district rules require a confirmation election for its board or other officials, a management plan must be submitted no later than three years after the date of the election. The management plans include a full accounting of the district's water demands and supplies, the resultant water need (shortage or surplus) within each district as well as the rules of the district. Each district's management plan is also required to establish a set of goals that the district will use to manage its water resources in order to meet its reported shortage or maintain a surplus water budget. GCDs are mandated to produce management plans during their initiation, as well as periodic updates over time. A GCD is required to reevaluate and readopt its management plan every five years. A district can resubmit its plan with or without changes, but the new submittal must be reapproved by the TWDB. Districts may their management plans more frequently than every five years, but it is uncommon unless there are major changes to the district's structure, resources or available data. If a district wishes to amend its management plan for updates of components other than the Modeled

Available Groundwater (MAG) or DFC, it must submit the proposed amendment to the TWDB so that it can determine whether the amendment requires formal approval or not. Any and all revisions to DFC and/or MAG automatically require TWDB approval. If a management plan is not approved, the denied district has two options: it can either appeal the denial to TWDB or revise and resubmit its plan within 180 days. Management plan denials are very rare and the TWDB has taken several steps in recent years to help districts craft approvable plans, such as trainings and detailed guidance documentation. If requested, TWDB offers technical assistance with the management plan development process, such as preliminary reviews and comments/recommendations prior to adoption. Permitting structure and permit fees are discussed in every management plan but are a bit different for every district. State law allows districts great discretion in the deployment of permit, production and transport fees collected by the district, but does set some limitations. Except in some cases, a GCD is prohibited from using revenues obtained from the transport fee to prohibit the transfer of groundwater outside of a GCD but is not prohibited from using revenues for paying expenses related to enforcement of Chapter 36 or the GCD's rules.

CHAPTER 2: LITERATURE REVIEW

2.1 Groundwater Management Approaches

Groundwater management strategies can be classified into three general types of approaches; water supply management, water demand management and integrated water resource management (Edalat, Abdi, 2018). Water supply management focuses on discovering and developing new sources of water supply. This type of thinking traditionally dominated the water policy decisions of the twentieth century because the aim was to confront water scarcity. To this end, new sources of water were more practical and economical than conveyance across great distances, for example from a preexisting water source to a new population center (Bithas, Stoforos, 2006). These general types of management styles can be further refined by several factors, including but not limited to; urban vs rural, arid vs humid, artificial recharge vs natural recharge. Water demand management emerged in the latter half of the twentieth century along due to technological advancements in conveyance and increasing resource scarcity. This shift from supply management to demand management steered several areas of the world towards sustainable processes of using of water through a focus on the consumptive behavior of water users and increasing interest in efficiency. Reduction and conservation started to become water management solutions, replacing the previous mindset of new source discovery (Edalat, Abdi, 2018). Water demand management strategies can fall into several different categories. The most common types of strategies are; water loss prevention, conservation, reuse, intermittent supply practices, market pricing and water metering. The final type of groundwater management strategy began to take hold in the 1970's along with the greater environmental movement in the US and abroad. This type of management is called integrated water resource management (IWRM) and can be described as holistic approach that encompasses both supply and demand, and attempts to account for externalities. The definition of IWRM according to Global Water Partnership (GWP) is

> a process which pro-motes the coordinated development and management of water, land and related resources, in order to maximize the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems.

One key component of IWRM is the development of coordinated plans that involve the consideration and input of all stakeholder groups. We can see echoes of this type of carefully considered management approach in the GCD groundwater management plans that are the subject of this research. The underlying movement of IWRM is shifting groundwater management from an individual well or wellfield to entire aquifer systems. In this sense, Texas' system of GCDs represents a half-step into this approach. This is because although there are a contingent of GCD boundaries that are drawn according to hydrogeological formation, the majority are based on county boundaries which causes the responsibly of managing the resources of a single aquifer to be split up among several parties.

2.2 International Groundwater Management

Globally, groundwater is increasingly significant due to its relative stability in terms of both quantity and quality, even in the face of evolving climate scenarios. In the international context, the European Union (EU) has championed the "Integrated System Approach" to groundwater management. The Water Framework Directive-2000 and subsequent Groundwater Directive-2006 (Villholth, et al, 2018) have allowed the EU to effectively harmonize its water and groundwater management approaches, despite great differences in infrastructure, hydrogeological formations and even political outlooks.

India, as a nation, is the largest user of groundwater on the planet. India uses more groundwater annually (251 km^3/year) than the second-most (China, 112 km^3/year) and third-most (USA, 112 km^3/year) global groundwater consumers combined (NGWA, 2018, see Figure 2 for a list of the top users. This is despite the fact that India is fourth internationally in annual groundwater recharge with only 432 km^3/year, behind the US (1383 km^3/year), China (828.8 km^3/year) and Russia (788 km^3/year) (Gun, Margat, 2013).

Figure 2: NGWA Global Groundwater Users Fact Sheet

Country	Population 2010 (in thousands)	Groundwater extraction			
		Estimated groundwater extraction 2010 (km^3/yr)	Breakdown by sector		
			Groundwater extraction for irrigation (%)	Groundwater extraction for domestic use (%)	Groundwater extraction for industry (%)
India	1224614	251.00	89	9	2
China	1341335	111.95	54	20	26
United States	310384	111.70	71	23	6
Pakistan	173593	64.82	94	6	0
Iran	73974	63.40	87	11	2
Bangladesh	148692	30.21	86	13	1
Mexico	113423	29.45	72	22	6
Saudi Arabia	27448	24.24	92	5	3
Indonesia	239871	14.93	2	93	5
Turkey	72752	13.22	60	32	8
Russia	142985	11.62	3	79	18
Syria	20411	11.29	90	5	5
Japan	126536	10.94	23	29	48
Thailand	69122	10.74	14	60	26
Italy	60551	10.40	67	23	10

Source: National Ground Water Association

India also has a long history of both excessive and highly destructive flooding during monsoon season and extended devastating drought during the dry season. Over time, Indians have developed ways to mitigate these massive swings in water availability through diversion and storage and mechanisms, such as dams. Recently, the country has begun to experiment with a process called Underground Taming of Floods for Irrigation (UTFI), which diverts flood waters towards targeted aquifer recharge zones. The water is collected and stored in the aquifers during the wet season for use in the dry season. This concept is similar to artificial recharge in that the aquifer is treated almost like surface reservoir storage, but the difference is that UTFI does not require pumping water into the ground. Instead of direct injection, UTFI utilizes "Percolation Tanks" to slowly drain the captured water into the aquifer in permeable areas and something called a "Recharge Shaft" in impermeable areas. Recharge Shafts are dug directly through the clay or silty bottom of water collection ponds in order to drain the surface water through the impermeable layers down into the aquifer. During the dry season, the stored water is retrieved via traditional groundwater wells (Saha, Dipankar, et al, 2018).

This new concept in India is especially enticing because it solves both the problem of flood and of drought, all while safeguarding the sustainability of the resource. The key to the success of this project has been integrated resource management. This includes universal buy-in from all stakeholders and bilateral communication between surface infrastructure/engineering and subsurface infrastructure/engineering groups that are typically separate.

Several other areas of the world have adopted successful integrated approaches such as the International Union for Conservation of Nature (IUCN) program in large parts of the Middle East, the International Water Management Institute (IWMI) initiatives across Africa and Asia and the non-ended Groundwater Management Advisory Team (GW-MATE) that was a co-production of the World Bank and the Global Water Partnership (GWP). The GW-MATE program ran from 2000-2010 and completed projects all over the world including the following countries; India, Brazil, Thailand, Yemen, Mexico, Argentina, Paraguay, China, Kenya, Nepal, Bangladesh and Venezuela (IGRAC, 2018). Although the GW-MATE projects were limited to a single country at a time, the methods were very much in line with IWRM. Each project began with a vulnerability analysis of the groundwater resource that spanned the total source capture area and included the implementation of system-wide coordination efforts. After the completion of the GW-MATE program, the Global Environment Facility (GEF)[2] with help from the World Bank, the FAO[3], UNESCO[4] and the IAH[5] created the Global Groundwater Governance Framework-for-Action with the goal of addressing global concerns over the unsustainable usage of groundwater and the increasing degradation of aquifers through cooperative planning, regional coordination and international best practices. By contrast, in the United States, federal acts such as CERCLA[6], accompanied by individual Superfund projects, tend to narrow the high-level focus to individual site planning. This site-specific approach results in the bulk of groundwater management policy left to the individual states.

2.3 Groundwater Management in the United States

In the United States, groundwater is the primary source of potable water for roughly half of the total population (Megdal, et al, 2018). At the federal level there are mandatory minimum water quality levels and discharge regulations on drinking water sources. However, although these federal limits exist, the majority of the laws,

---

[2] Global Environment Facility https://www.thegef.org/

[3] The Food and Agriculture Organization of the United Nations is a specialized agency of the United Nations that leads international efforts to defeat hunger. http://www.fao.org/home/en/

[4] The United Nations Educational, Scientific and Cultural Organization

[5] International Association of Hydrogeologists https://iah.org/

[6] Comprehensive Environmental Response, Compensation, and Liability Act, known also as Superfund

governance and resource management is left to the states. Some states regulate groundwater at the state house, others delegate the responsibility to even smaller jurisdictions like the GCD system of Texas. The way that groundwater is managed from state-to-state depends greatly on the level of human reliance on groundwater which is very variable. For example, Virginia relies on groundwater for less than 5% of its total water supply, whereas Kansas' water supply is composed of approximately 80% groundwater. Because of this great variance in groundwater reliance, state-to-state management approaches and planning priorities are also highly varied. Even the legal framework surrounding groundwater spans a wide range, many states do not acknowledge the connection between groundwater and surface water supplies when dealing with property rights and source determination issues (Megdal, et al, 2018). Surprisingly, geography and even physical formations types do not tend to define the boundaries between groundwater management approaches. Instead, sustainable groundwater management approaches tend to be adopted more in accordance with the political composition of state legislatures.

Some states and institutions have taken steps to evaluate their groundwater management policies by analyzing aspects of different types of financial apparatuses on groundwater flow balance or recharge rates. California has examined transaction costs and pumping constraints data to determine what are the user and resource characteristics that determine the success of collective remedy implementation (Ayres et al, 2018). A study in Colorado looks at conservation measures through examining the impacts of self-imposed groundwater pumping fees on groundwater use (Smith et al, 2017). A study in Kansas demonstrated the effects of groundwater management on land values by comparing counties where groundwater management is active to unmanaged counties (Edwards, 2016). All three of these studies have shown that economic incentives or penalties can be effective in preventing excessive withdrawal and encouraging sustainable use practices. However, these studies also discuss how difficulties arise in developing a true statistical test due to the diversity in the managing entities that covering one aquifer. This could indicate that the best management of a single common-pool resource like groundwater is done at a higher (statewide, e.g.) level.

It can be said that the most obstructive aspect of groundwater management in the United States is the disambiguated nature of its groundwater governance. Beyond the lack of coordination between individual states, the separation of surface water governance and groundwater governance tends to be ineffective due to the lack of consideration of all stakeholders in the system which can lead to overuse and under-planning. In this type of governance, the disconnected management framework exacerbates existing overexploitation of the common-pool resources because individual governing groups are too segmented to take action and lack the empowerment of critical evaluation over the system as a whole (FAO, 2016). A study from 2012 examined the impacts of different types of management practices on groundwater drawdown as a response variable. That paper proposes that one way to combat open plunder of common-pool resources like groundwater is to determine the most sustainable method of resource management through quantitative evaluation (Madani and Dinar, 2012). This project proposes a similar method of evaluation in order to better understand the efficacy of Texas' groundwater management structure and empower stakeholders to make informed decisions.

2.4 Literature Review Summary

Groundwater availability, and groundwater management strategies can be modeled and evaluated, there are even several methods for optimal groundwater use simulations (Wagner and Gannett, 2014), however, few if any techniques exist to evaluate management policy. This is a problem because resource managers could at once have access to the best possible groundwater availability models and at the same time have no way of knowing if the policy of their district is helping or hindering their strategy. There are several approaches to applying groundwater modeling to groundwater management but there are also so many ways to craft the policy discussion that it is difficult to determine the most impactful approach.

There has been, and will continue to be, groundbreaking work done with groundwater management on a national and international scale. However, the one component that many of these projects and initiatives lack are analytical tools to quantifiably measure impacts. In some cases, it seems that the missing element to support the claims of efficacy for one groundwater management strategy over another is hard data. There are areas where that is not the case, like in the realm of understanding the economic impacts of groundwater policy decisions. However, for areas like Texas where property owners have the ultimate right to the resource instead of the governing body, GCDs are not as empowered to implement these types of financial incentives or penalties. Several studies have analyzed the impacts of different management strategies to determine the most impactful or sustainable. In Texas the management structure is largely homogenized due to all-district required goals. This paper does not endorse one management strategy over another, but what it does do is propose a method by which Texas' groundwater policy can be examined.

CHAPTER 3: DATA MINING

3.1 Data Mining Methodology and Issues Encountered

3.1.1 DATA MINING CONTENT DEVELOPMENT

The data mining section of the project began with an accumulation of all available groundwater management plans from the TWDB website. They were downloaded one-by-one and then catalogued in an internal folder structure by GCD, in Portable Document Format (PDF). The page lengths ranged from less than ten to over 400 and most were a mix of text, images and tables. The older PDFs often contained the full set of rules for the district and several of the documents also included clippings from local newspapers covering the districts and their educational outreach events, which helped GCDs proves they had satisfied a state-mandated public outreach and engagement goal[7]. These newspaper clipping sections were difficult to work with because they were often scanned using old copier equipment, resulting in low-resolution capture. This made the text barely legible to the naked eye and even less decipherable to the computer-aided text recognition programs. There was the added issue of whether to include the newspaper content as part of the final data or not, as the clippings were often full pages of a newspaper where only one small section referred to the groundwater conservation district in any way. The decision was made to exclude the newspaper pages from the final data results as most of the scanned pages were just creating noise in the data and often the whole excerpt that pertained to the district was captured elsewhere in the document under the district's goals section.

[7] A table of the complete goal and information requirements for management plans is shown in chapter 1, section 4, subsection 3 of this report.

3.1.2 DATA DEVELOPMENT

An open-source extension in the CRAN (Comprehensive R Archive Network)[8] repository for R studio called PDFTools[9] was used to create readable content from the GCD Management Plan PDF documents. The PDFs were downloaded directly from the Texas Water Development Board Website and then run through the PDFtools package to create searchable text in a table format. Some PDFs, mostly much older ones, required manual cleanup of the PDFtools output, see Figure 3 for an example.

Figure 3: Example of raw processed text from older Management Plan, prior to manual correction.

Drainage of Crockett County is by means of:intermittent, dendritic streams. On the east side of the county a dry tributary ofDevils River drains southeastward into Sutton County. Johnsons Run and owards, re k bi,sect:ceiitral (:l'.o(?kett County-an drajn; U: ward, jo?,Iiing-Devils,River.:an4 the Pecos Riv., respectively; m:Vai Verde Cotilify. Jn,ihe Nt>rthwestetn pa fotCtockett County, Live-'" Oak Creek drains southward :into the Pecos River at Lancaster Hill. The dry bed of Spring Creek origitt.at s inthe p..ortlteastern cotner of the county and runs northeastward Generally, t.lte county can be said.to e in:thRioGrat1(le-4ni,itagebasin. Only the extreme nortli\:}astem comer c.fthe cpunty lies in the 'Colorado River· drunage basin/

Groundwater Reso c.es of the,Emerald u.W.C.D.

The primary sources of groundwater in Crockett County are derived from the Edwards-Georgetown aquifer of Cretaceous age, sands of the Trinity Group or Trinity aquifel'. and unconsolidated alluvium of Quaternary age which overlies the older Cretaceous rocks principally along the Pecos River, Live Oak Creek, Howard Creek and Johnson Draw.

This type of manual cleanup was not required for any of the PDF documents that were produced after the year 2005, which includes the bulk of the management plans. From the processed PDF information, a table of word counts was developed, searching for pre-determined keywords. All of the counts were then assembled in a master table with one observation entry for each processed document. The final list of keyword

[8] https://cran.r-project.org/

[9] https://cran.r-project.org/web/packages/pdftools/pdftools.pdf

counts, the additional document metrics and the response variables can be found in Table 3. The variables generated from the data-mining outputs are recorded as “Count of Keyword”, all other variables were captured directly from the management plans.

Table 3: Variable Types, Descriptions and Ranges

Variable Type	Variable Name	Description	Range
Predictor	District Count	Count of "Keyword"	34 to 1964
Predictor	Storage Count	Count of "Keyword"	0 to 172
Predictor	Permit Count	Count of "Keyword"	1 to 605
Predictor	Management Count	Count of "Keyword"	5 to 924
Predictor	Aquifer Count	Count of "Keyword"	0 to 856
Predictor	Recharge Count	Count of "Keyword"	3 to 257
Predictor	Withdraw Count	Count of "Keyword"	0 to 177
Predictor	Shortage Count	Count of "Keyword"	0 to 76
Predictor	Conservation Count	Count of "Keyword"	4 to 361
Predictor	Resource Count	Count of "Keyword"	1 to 136
Predictor	Groundwater Count	Count of "Keyword"	21 to 959
Predictor	Precipitation Count	Count of "Keyword"	0 to 201
Predictor	Supply Count	Count of "Keyword"	2 to 203
Predictor	Demand Count	Count of "Keyword"	1 to 164
Predictor	Words Per Page	Document Words per Page	54 to 397
Predictor	Total Word Count	Total Document Word Count	1621 to 143,266
Predictor	Pages	Pages in Document	6 to 421
Predictor	GCDname	Name of GCD	Bandera County River Authority and Groundwater District to Wintergarden GCD
Predictor	Year	Publish year of Document	1998 to 2018
Response	Water Supply Need Total	Total Water Supply Need in acre-feet	-971,421 to 1,448,676
Response	Water Need Binary	Water Supply Need Classification	"1","0"
Response	Water Need	Water Supply Need Classification	"shortage", "surplus"

The most difficult data to assemble for this dataset were the response variables, which ended up being a TWDB-reported value called ‘water supply need’. There are three response variables because, although they all record the same measurement (Water Supply Need), it is expressed in different ways for use with different model types. The variable “Water Supply Need Total” is the total recorded water supply need in acre-feet. The variable “Water Need” is a simplification of the reported number in acre-feet to either ‘shortage’ if the water supply need is negative or ‘surplus’ if the water supply need is zero or positive. The variable ‘Water Supply Need Binary’ is a variation of “Water Need” where ‘shortage’ is expressed as a “1” and ‘surplus’ is expressed as a “0” for use in binary response models. Water supply needs are defined by the Texas Water Development Board as “…projected water demands in excess of existing supplies that would be legally and physically available during a drought of record” (TWDB, 2015), the plan goes further to elaborate:

> Although in some regions it appears that there are sufficient existing water supplies region-wide to meet demands under drought conditions in the early planning decades, local existing water supplies are not always available to all users throughout the region. Therefore, water needs were identified as a result of this geographic “mismatch” of existing supplies and anticipated shortages
>
> And
>
> When existing water supplies available to a specific water user group are less than projected demands, there is a need for water. In other words, once there is an identified water demand projection for a given water user group, this estimate is then deducted from identified existing supplies for that water user group, resulting in either a water supply surplus or a need. (TWDB, 2017)

It was difficult to select a response variable for the modeling component because in order for the results to be useful, the response had to be either an integer or a metric that was conducive to binary expression and had to be common to all GCDs, most plans, and most plan years. Water Supply Need was chosen as the response variable not because it is a perfect measure of the storage/surplus groundwater availability within the districts but because it is the closest fair approximation of this type of metric. It was quickly apparent that attempting to produce this metric from GAM runs was going to be impossible because of the inconsistencies in the method of apportioning MAG between districts because there has never been a need nor impetus for standardization across Groundwater Management Areas. Additionally, some management planning areas have been slower to develop complete GAM run reports and there is further inconsistency in adoption of GAM results in GCD management plans. Ultimately, Water Supply Need emerged as the most consistent and quantifiable metric to use as a response variable for the models. When the year of the management plan fell between years with reported water supply need data, the two values were averaged together to approximate water supply need for the time of the management plan creation. For example, if a district's management plan year was 2015 and the water supply needs were only reported for 2010 and 2020, those two values were averaged together to generate an approximate water supply need for 2015. Although Water Supply Need as it is reported in the state water plans is an accounting of both the groundwater and surface water shortage/surplus, the vast majority of proposed water management strategies proposed to address these needs fall into two categories 1. Conservation and 2. Expansion of groundwater use, where 99.8% of the total volume is groundwater use.

Insights from the data mining process were used in order to develop the predictor variables for the dataset. A short list of Variable keyword selection was assembled from commonalities among the planning documents to ensure even distribution. Next, content mining methods were employed to pull out a handful of highly repeated words (district, groundwater, resource, conservation), and then the rest of the keywords were chosen for their potential to be relevant signifiers. For example, shortage was chosen as an attempt to identify districts that are using their management policies as tools to resolve drought or water shortage issues. The word 'storage' was selected as a nod to artificial storage and recovery projects, and also as an experiment to see if there is a difference between districts with higher/lower overall reserves and how they describe their resource availability. Precipitation was chosen to test for a connection between instances of the word precipitation and the word recharge and also to figure out if there is a geographical difference between districts that refer to precipitation and those that ignore it in their management plans. The word 'withdraw' was initially captured as 'withdrawal' but was later recast in order to capture both 'withdrawal' and 'withdrawn'. The word 'management' was chosen over 'manage' to separate the concept from the action as a way of approximating the number of instances that the plans are self-referential. The word 'permit' was selected to capture permitted, permitting, and permit as the contextual occurrence of these terms are not meant to be limited by their expression. In all cases the contextual subject would be wells and therefore there is less concern that erroneous definitions would pervade the model as with manage versus management, as manage is a more common term. When representing water supply need as a binary variable, 0 was chosen for surplus and 1 was chosen for a shortage in all models.

3.1.3 FINAL OUTPUT DATA

When all of the management plan documents were through the data mining process, a master spreadsheet was assembled with one tab per plan, organized first alphabetically and then by plan year. The next challenge was to assemble each GCD plan (tab information) into a single line of data (observation). Several iterations of data grabs were pulled from the sheets, partially through the use of excel equations and partially through manual adjustment. Once all the keyword counts were pulled out into a single table, the following additional information was assigned to each observation:

- GCD Name
- Document Page Count
- Document Word Count
- Document Words Per Page (Page count divided by Words per Page)
- Plan Year
- Water Supply Need Total

As noted previously, the final response variables were 'Water Supply Need Total' (actual values in acre-feet), "Water Need" (classes of 'shortage' and 'surplus') and 'Water Supply Need Binary' (water supply classes represented as "1" or "0"). In order to populate the binary response variable, if the water supply need was negative it was set to "Shortage" and if the water supply need was zero or positive it was set to "Surplus". Water supply need is a value developed by the TWDB and it is expressed converse to logic, where they represent a water need as negative. This is confusing because if a GCD has a *negative* amount of *Supply Need*, it would follow that that actually indicates a lack of need and therefore a water surplus. But the TWDB chose to record their water supply need values differently so it is important to note here that the project data also reflects this terminology choice.

Before building the series of models I wanted to inspect the aggregated data mining results to see if there were any obvious patterns in the dataset. A few initial observations when working with the plan documents were how much the plans grew in terms of page length over the years and as the documents became longer the word counts did not grow in proportion. As shown in Figure 4, the total number of pages in the document is much higher in the more recent plans. There are a handful of early plans that are less than 20 pages long, whereas the documents published after 2015 are almost all around 100 or more pages in length. I expected to see a much more extreme growth in pages per year than Figure 4 shows because all of the longest documents were developed after the year 2010. However, as depicted in Figure 4 there is still a notable number of documents that are less than 100 pages long even into the year 2015. However, at the same time word counts level off over time, in fact the highest word densities belong to plan documents that were published between 2000 and 2010 as shown in Figure 5. Unlike the total page length, the total words per page remain relatively steady over the full span of publication years, as shown in Figure 5. There is a gradual transition to fewer words per page over the full span of the publication years. The shift can probably be attributed to the more recent move towards including several tables that take up the space of several pages but do not contain much text. The majority of these tables are excerpts from the TWDB state water plan. The addition of these tables may be in response to new goal requirements for GCD management plans, such as addressing Desired Future Conditions. This type of goal may extend the overall length of the documents while shortening the total word count because they often require lengthy tables of data from GAM runs in order to enunciate the goal and the achievement metrics. As discussed earlier in Chapter 1: Project Background, the new goals can be viewed as an attempt to more closely align the individual GCD management plans to the overall state water management plan. One example of that alignment is that GAM runs are now included in all GMA management plans as well as the TWDB state water plan.

These types of patterns in the plan documents are important to note before constructing the models because distinct trends in one, or a few, variables that are directly correlated to time period (plan year) may impact the other variables. Most of the assembled variables are keyword counts, and if the number of total words is directly impacted by the plan year, then it would follow that the word counts would be skewed over time as well. In order to account for this type of inherent correlation between the variables I developed a fixed effects model to control for the “Year” variable. This will be discussed more in Chapter 4: Model Methodology.

Figure 4: Plot of pages by year

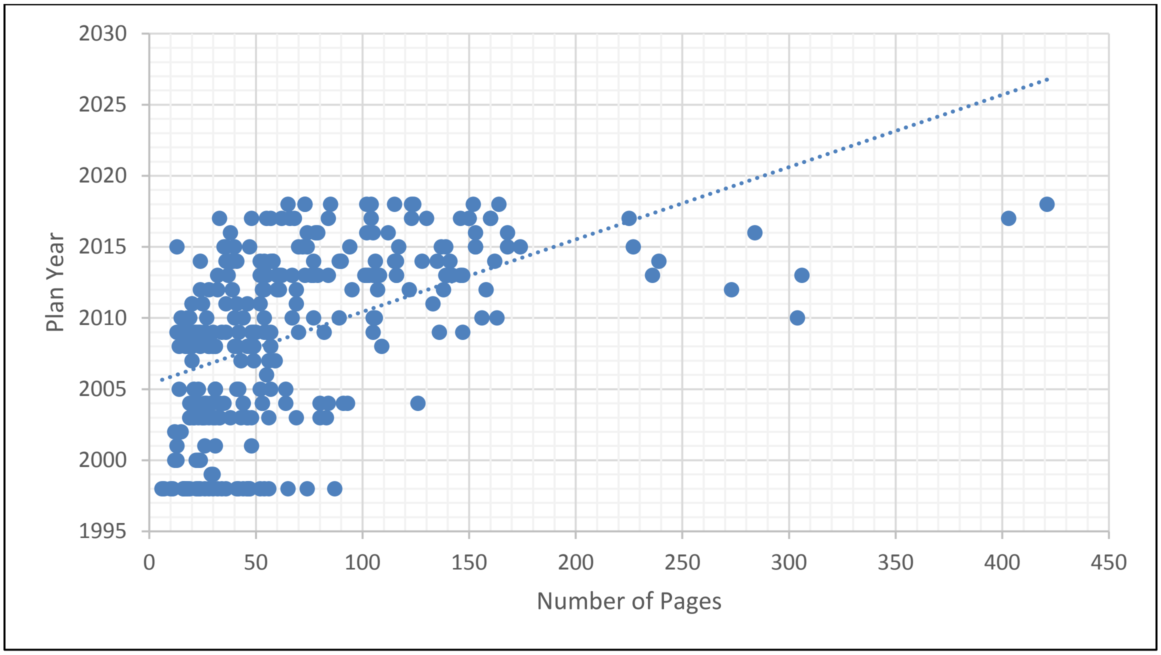
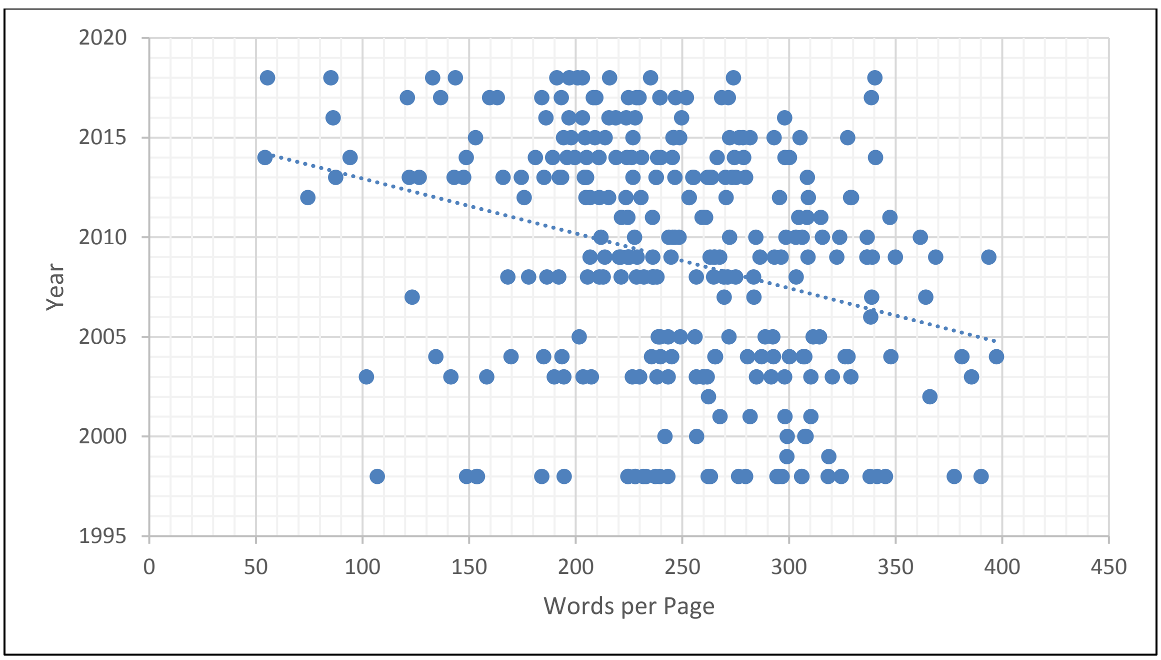
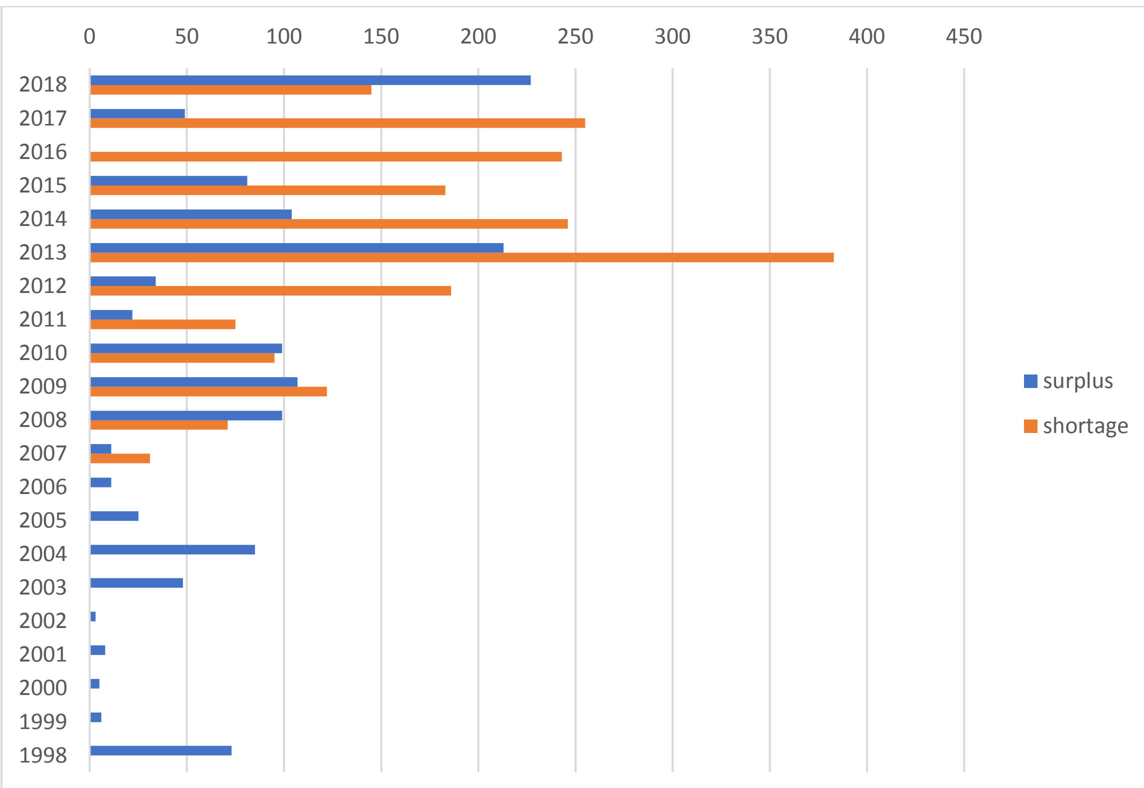


Figure 5: Plot of words per page by year



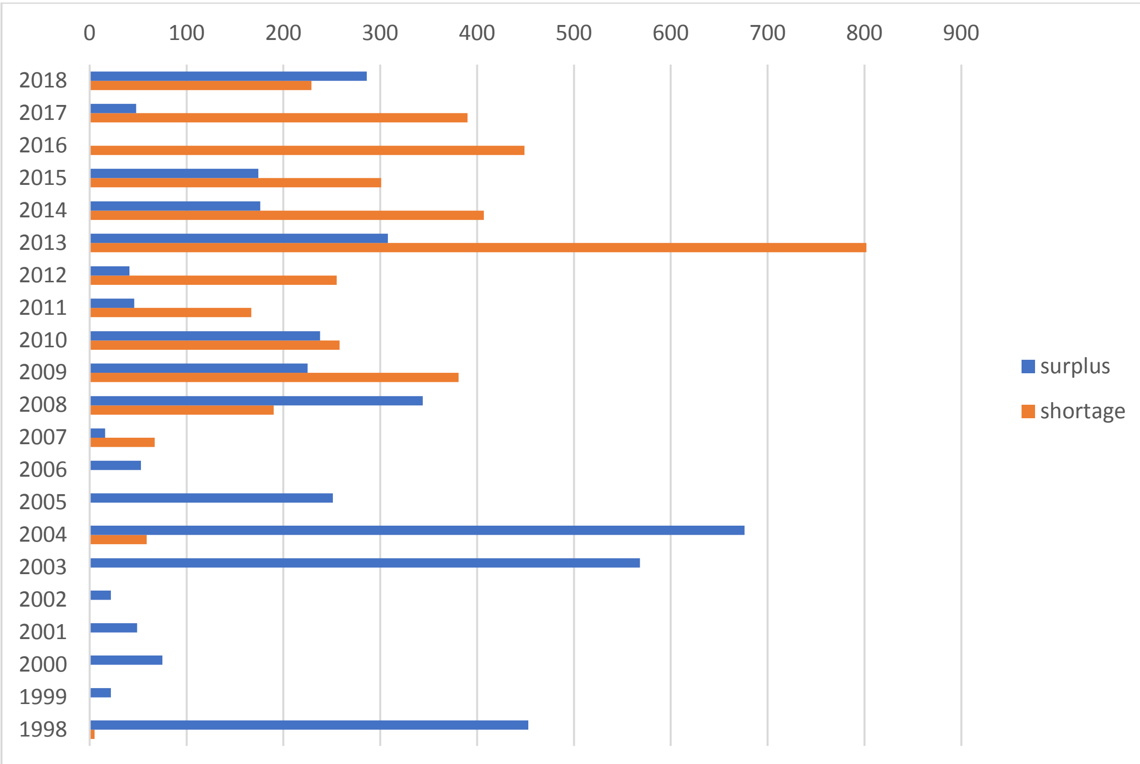
The raw data displayed several patterns that are more thoroughly fleshed out in the subsequent model sections. Figure 6 displays the count of Precipitation by Water Supply Need class and Plan Year. As we can see, the years that show the highest number of Precipitation count also roughly coincide with the shortage class up until the year 2018. Another interesting shift in this figure is the growth trend of Precipitation count across both classes of plan documents that follows the 2011 drought.

Figure 6: Precipitation count by Water Supply Need Class and Plan Year



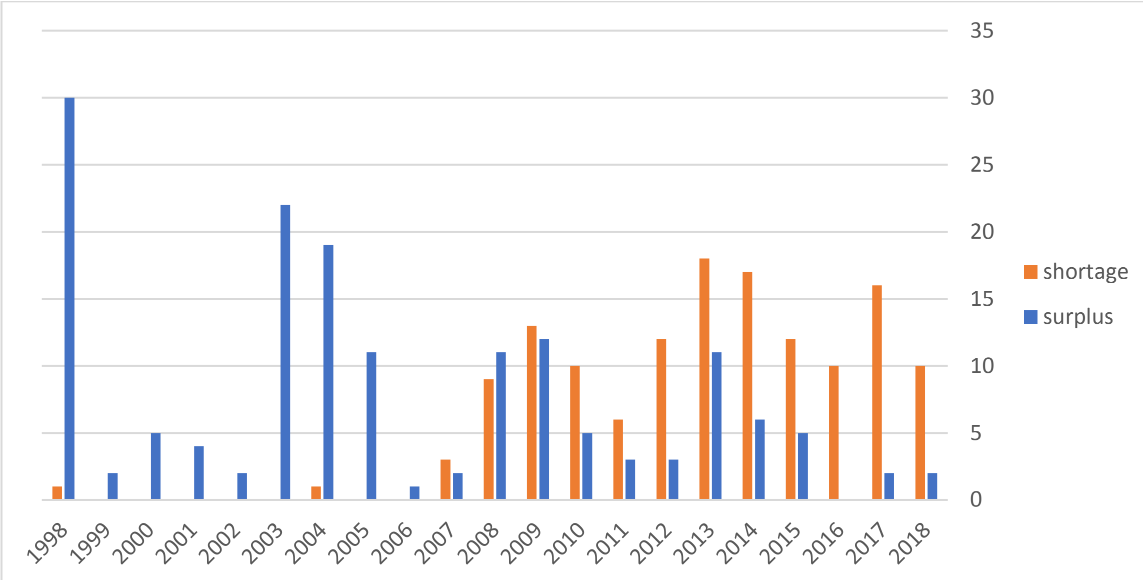
In the same vein, Figure 7 displays the count of Recharge by Water Supply Need class and Plan Year. Here we would expect to see the highest counts for Recharge belonging to the years that display a Surplus class but as it turns out, the Recharge count is relatively balanced between the classes.

Figure 7: Recharge count by Water Supply Need Class and Plan Year



Another item to consider in the raw data is that overall the number of GCDs that reported a water supply shortage was very low until 2007. Figure 8 displays the plan documents by Year and Water Supply Need class and paints a vivid picture of the progression of resource drawdown across all the GCDs over time. After 2008 the number of plans that report a water supply need surplus are always less than those of shortage. Especially in 2016 to 2018, surplus-class GCDs are in the significant minority of the plans published in those three years.

Figure 8: Chart of plan documents by Water Supply Need Class and Plan Year



On the next two pages we can see the same charts as Figure 6 and Figure 7 but instead of displaying the data by plan Year, it is by GCD name. Unfortunately, due to page size constraints it's impossible to show every single one of the 100 GCDs. Figure 9 shows how across the GCDs, a higher count of Precipitation is typically indicative of the Shortage water supply class. The chart does show Goliad GCD as a significant outlier to this overall pattern, but every GCD is different so that is to be expected. Again in Figure 10 we can see Hemphill GCD as an outlier, but generally the higher counts of Recharge indicate GCDs that fall in the Surplus class. For a complete breakdown of all the predictor variable keyword counts by both GCD and Year refer to the charts provided in Appendix A.

Figure 9: Precipitation count by Water Supply Need Class and GCD Name

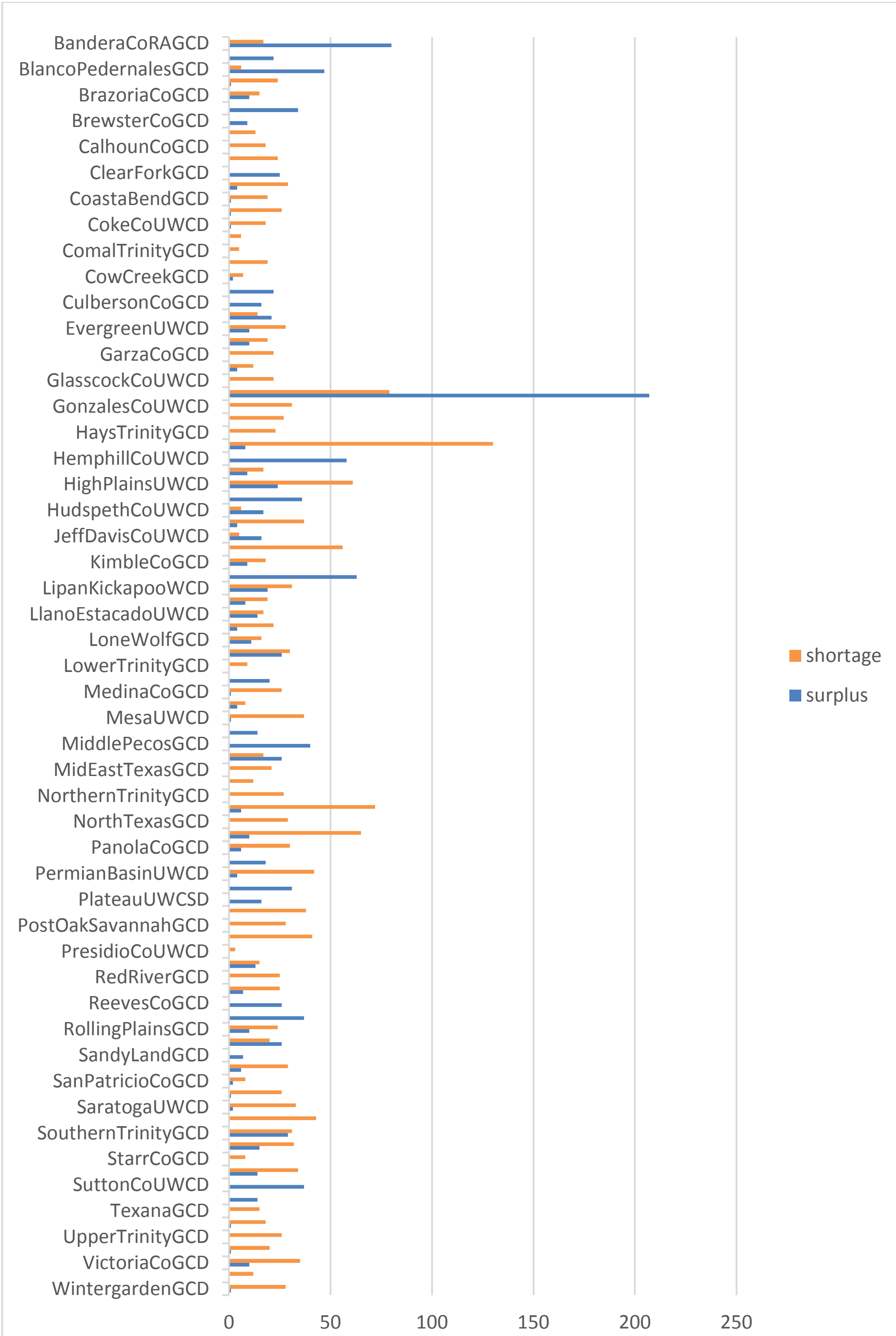
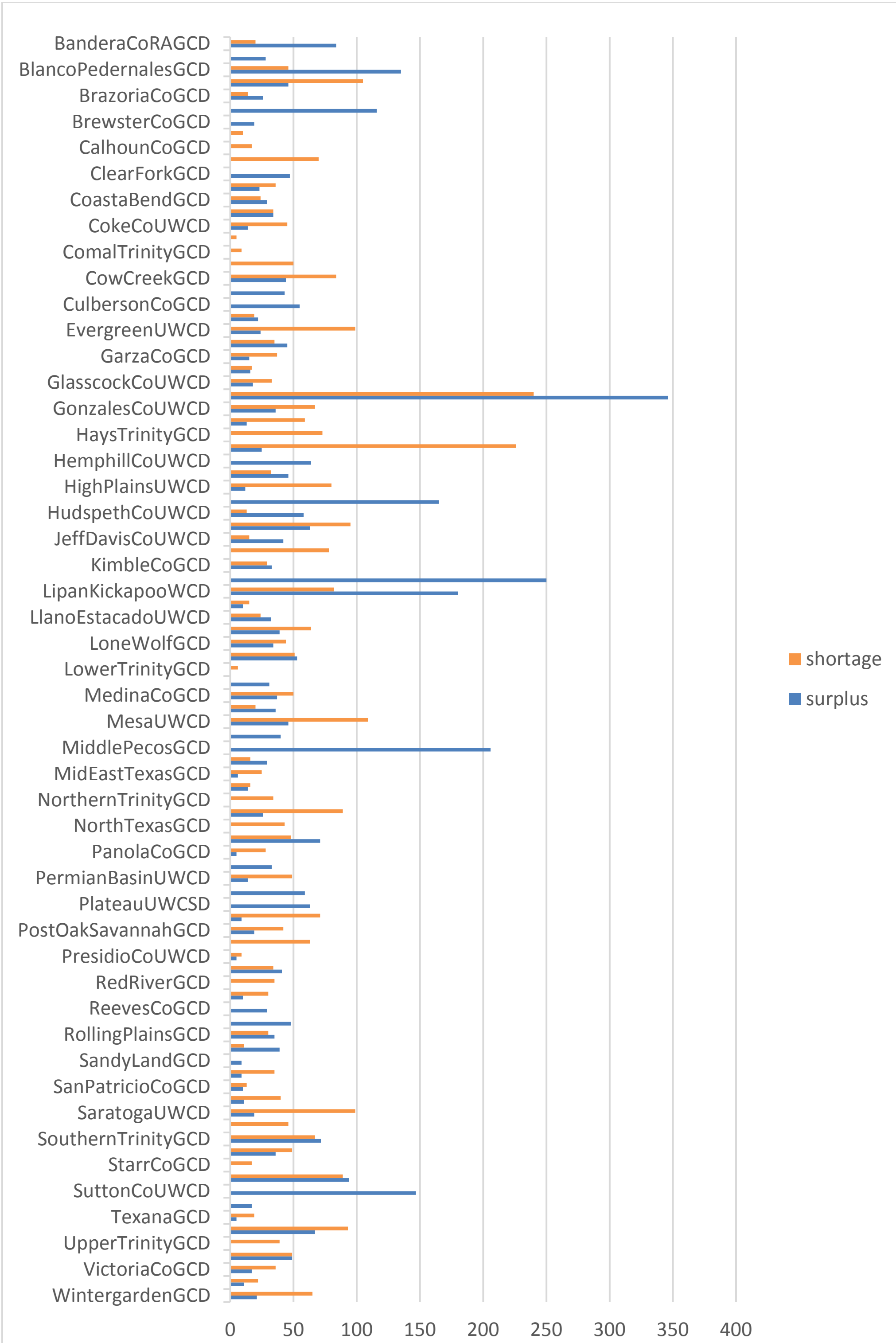


Figure 10: Recharge count by Water Supply Need Class and GCD Name



CHAPTER 4: MODEL METHODOLOGY

4.1 Model Introduction

Before we begin the discussion of model development let's return to the stated goals from Chapter 1:

1. Determine whether GCD management plans can be used to classify GCDs by water supply need.
2. Discover what components have the strongest impact on classification.

For the first goal, in order to evaluate the impact of the GCD management plans on the GCDs the models will use the predictor variables to classify the GCDs into their categories of water supply shortage or water supply surplus. The next goal will be achieved by interpreting the model results to determine which predictor variables have the strongest relationship with the response variable (Water Supply Need).

4.2 Statistical Model Development and Discussion

4.2.1 LINEAR REGRESSION MODEL DEVELOPMENT

The first model to be built was the simple linear regression model. The formula for the linear regression model can be found in Equation 1.

Equation 1: Linear Regression Model Equation

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots$$

where Xj represents the jth predictor and βj quantifies the association between that variable and the response. We interpret βj as the average effect on Y of a one unit increase in Xj, holding all other predictors fixed. (James, et al, 2017)

Where Y represents the response variable, which is "Water Supply Need Total" in this model. The numbered betas represent the coefficient estimates and the set of (X_1, X_2…Xn) where each X_n represents a predictor variable from the dataset (i.e. a value of "Resource Count" or Words per Page).

There can be as many as six different plan iterations for a single GCD, and numerous plans share the same publication year. Thus, plans from the same GCD are correlated across years, and some common intertemporal shocks (for example, a statewide drought) will affect all plans across GCDs within the same year. In order to account for this panel structure in the data, three fixed effects models were constructed. The equation for the fixed effects model is Equation 2 where Y_{it} represents the response variable, for this model that would be the variable "Water Supply Need Total" as it was in the original linear regression model.

Equation 2: Fixed Effects Model equation

$$Y_{it} = \mu_t + \beta X_{it} + \beta_2 X_{2it} \ldots + \gamma_i + \alpha_t + \varepsilon_{it}$$

The set of predictor variables are represented by X_{nit} (X_{1it}, X_{2i}t, etc.) these are the same as our independent variables from the linear regression model ("Conservation Count", "Words Per Page", etc.). The set of GCD fixed effects are represented by γ_i which in this case would be "GCD Name". The parameter α_t is a "Year" fixed effect.

4.2.2 Logistic Model Development

In the logistic regression, we are trying to predict the probability that the value of Y (the response variable) which can equal 0 or 1, is equal to 1. Each class is represented as either 0 or 1, in our case an observation that records a water supply surplus has a value equal to 1 and an observation that records a water supply shortage has Y= 0. Unlike in the linear regression, the coefficients in the logistic regression model are not interpreted as marginal effects, which must be calculated afterward.

4.2.3 Linear Discriminant Analysis Model Development

Linear Discriminant Analysis (LDA) models focus on maximizing the separability between the response variable classes. LDA models are similar to Logistic Regression, but instead it models the distribution of the predictors (X_1, X_2…) separately for each of the classes of the response variable (Y) and then uses Bayes' theorem to flip the estimates for Pr(Y = k|X =x) (James, et al, 2017).

Equation 3: Bayes Theorem, James, et al, 2017

$$\Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)}$$

The LDA classifier works by assuming that the data points within each class come from a normal distribution with a class-specific mean vector and a common variance and then estimates for these parameters are plugged into the Bayes classifier (James, et al, 2017).

The LDA models are validated through a technique called leave-one-out cross-validation. K-fold cross validation is a resampling method that is used to evaluate machine learning models on a data sample. The k-fold cross validation process uses a single parameter (k) that indicates the number of groups into which the sample data is split, i.e. if k=10 the data is split into 10 groups. Leave-one-out cross validation is a variation of k-fold cross validation where K is always equal to N (the number of observations). In this way the model is trained on all of the data except for one observation and the error is evaluated by iterating through the entire dataset, leaving out one point at a time. A corresponding fixed effects model type does not exist for LDA so Year and GCD Name predictor variables were left out of the model. Please note that due to this omission these model results cannot be evenly compared to the results of the Linear Regression model.

4.2.4 Support Vector Machine Model Development

The concept of Support Vector Machines (SVM) first came about in the 1990's in the field of computer science as a way to resolve classification problems (James, et al, 2017). Three Support Vector Machine (SVM) models were constructed to classify the data into two classes (Water Supply Need = Surplus, Water Supply Need = Shortage).We know from the other classification models (LDA, logistic regression) that there are certain predictor variables (keyword counts for Precipitation, Conservation, etc.) that can be used to classify the GCDs into the two classes. However, the difference is that LDA is an analytical solution and SVM is an optimization solution. Where LDA is using the entire dataset to estimate covariance matrices, SVM models are optimized to focus on the subset of the data that lies within the separating margin.

SVM is a generalization of the Maximal Margin Classifier (MMC) technique. It is a classification approach that enables non-linear decision boundaries between the classes. Each observation is sorted to one side of the other of the margin, creating a binary classification. An MMC is a hyperplane that separates the data into two classes with the widest margin (James, et al, 2017). Although the MMC can only classify data that is linearly-separable, the support vector classifier uses the kernel trick to handle data that does not have a linear decision boundary. One of the primary advantages of SVM classification over LDA or logistic regression is that because the margin allows a set budget of crossed (incorrect) observations it is less sensitive to outliers. The way that SVM models provide insight into the relative difficulty or ease of classification within a dataset is through the amount of flexibility in order to establish the marginal boundary and its width. The way to compute which side of the decision boundary an observation falls (to which class the observation belongs).

Equation 4: Equations for determining vector classes using the Decision Rule, James, et al, 2017

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} > 0 \text{ if } y_i = 1,$$

and

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} < 0 \text{ if } y_i = -1$$

so that the equation for the hyperplane is

$$y_i (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}) > 0$$

which means that we can classify a test observation by plugging it into Equation 4 and solving to determine if the outcome is positive or negative. The magnitude of x also indicates the distance from the hyperplane, large values are points that are well within their class assignment and smaller values are points closer to the decision boundary. Linear SVM models also make use of a tuning parameter for optimization (cost) so in order to do that in the linear kernel model we take Equation 4 and add slack variables.

Equation 5: SVM Linear Kernel Tuning Parameter Equation, James, et al, 2017

$$y_i\,(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \geq M\,(1 - \varepsilon_i)$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^{n} \epsilon_i \leq C$$

Where M represents the width of the margin and C is a nonnegative tuning parameter and $\varepsilon_1 \ldots \varepsilon_n$ are the slack variables that allow a number of observations (set by C) to be placed on the wrong side of the hyperplane so that the classification accuracy of the model as a whole can be higher. In the case of polynomial kernel models there is an additional parameter, degree which is a positive integer that allows support vectors to be fit in a high-dimensional setting, rather than in the original feature space.

Equation 6: The equation for the Radial Kernel SVM model, James, et al, 2017

$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2).$$

Equation 6 is the formula for another non-linear kernel (radial kernel) that uses the tuning parameter gamma to force the model to consider the Euclidean distance between test observations and training observations. The radial kernel operates in a localized fashion so that only the nearby training observations impact the classification of a given test observation (James, et al, 2017).

CHAPTER 5: MODEL RESULTS

5.1 Regression Model Results

5.1.1 LINEAR REGRESSION MODEL

In the linear regression model the three estimates that are statistically significant at $p \leq .01$ are count of precipitation, conservation and groundwater, so we can reject the null hypothesis that the estimates for those variables are equal to zero. Each of these effects represents a statistical association, not a causal effect. The estimate for Precipitation Count is -3,415. That means that if precipitation count were to increase by 1, this is associated with an increase in the total water supply shortage[10] of 3,415 acre-feet, on average. The Conservation Count estimate is -882, so as conservation count increases by 1, water supply shortage would increase by 882 acre-feet on average. Similarly, with Groundwater Count, for every increase of 1, we would expect to see an increase in water supply of 459 acre-feet on average.

[10] Refer to the discussion of TWDB reporting logic for Water Supply Need recording vocabulary in Chapter 3, section 1, subsection 3 of this paper

Table 4: Linear Regression Results

Simple Linear Regression Results:					
Residuals:					
Min	**1Q**	**Median**	**3Q**	**Max**	
-907925	-21346	1976	24726	1408732	
Coefficients	**Estimate**	**Std. Error**	**t Value**	**Pr(>|t|)**	
(Intercept)	87461	54024	1.619	0.1066	
District Count	-208.19	115.81	-1.798	0.0733	.
Storage Count	-177.99	662.63	-0.269	0.78843	
Permit Count	-61.92	200.31	-0.309	0.75745	
Management Count	43.54	230.44	0.189	0.85027	
Aquifer Count	-52.91	118.28	-0.447	0.65498	
Recharge Count	416.70	396.67	1.05	0.29441	
Withdraw Count	-490.67	882.12	-0.556	0.5785	
Shortage Count	811.70	2230.22	0.364	0.71617	
Conservation Count	-881.86	283.35	-3.112	0.00205	**
Resource Count	341.57	932.80	0.366	0.71451	
Groundwater Count	458.69	152.70	3.004	0.00291	**
Precipitation Count	-3415.19	1033.30	-3.305	0.00107	**
Supply Count	-233.02	712.58	-0.327	0.74391	
Demand Count	-776.52	566.21	-1.371	0.17135	
Words Per Page	-265.90	209.89	-1.267	0.20627	
Total Wordcount	5.16	2.64	1.952	0.05196	.
Pages	-651.00	644.76	-1.01	0.31352	
Year	-3717.00	1907.00	-1.949	0.052274	
Residual standard error: 125500 on 278 degrees of freedom					
Multiple R-squared: 0.1188, Adjusted R-squared: 0.0649					
F-statistic: 2.204 on 17 and 278 DF, p-value: 0.004464					

The multiple R-squared value was 0.12 and the adjusted R-squared value was 0.06 for the linear regression model. The adjusted R-squared results take into account the number of predictor variables in the model, whereas the Multiple R-squared results do not. R-squared results are a type of goodness-of-fit metric for linear regression models. These values represent the cumulative percentage of variation in the dependent variable for which the response variables account. In general terms, R-squared is a measure of the strength of the relationship between the dependent and independent variables, in this case the model is showing a low (12%) percentage. This doesn't necessarily mean that the model outcomes are inadequate, but it does signify high variability in the model, which we can also observe in Figure 11.

Figure 11: Linear regression model Residuals vs Fitted plot

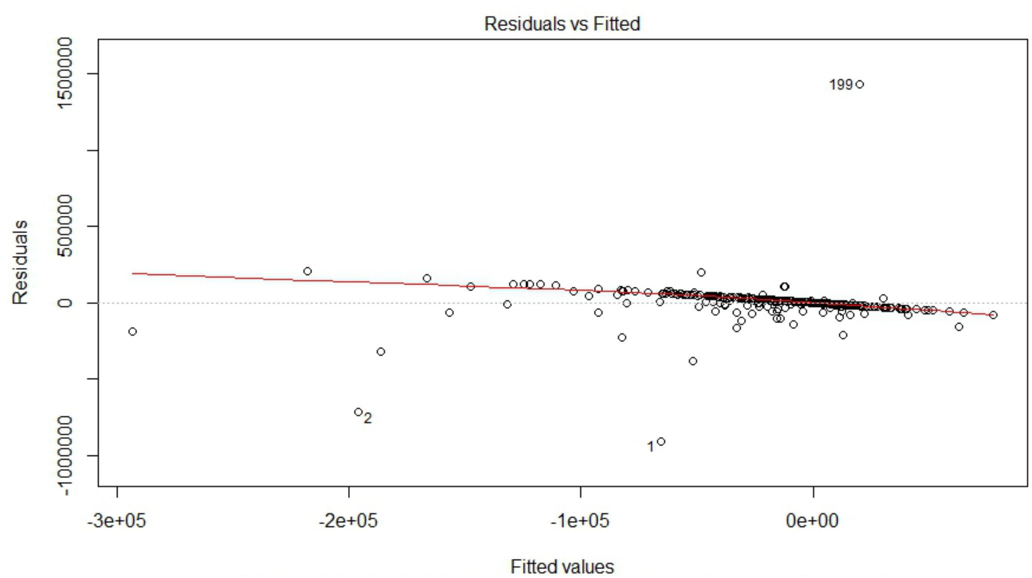
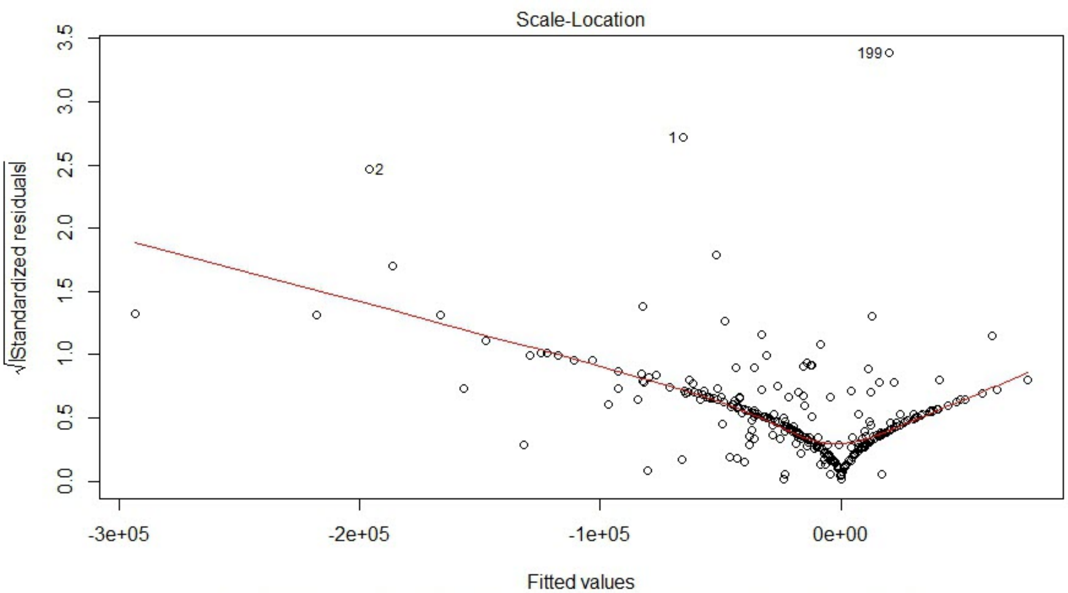


Figure 12: Linear regression model Scale-Location plot



The Residuals vs Fitted plot shows strong linearity and good model fit, but there are also a few significant high and low outliers. A similar trend is shown in Figure 12 where the overall model appears to have good fit but with obvious outliers (observations 199, 1, 2). Additionally, an examination of the error reporting section of the linear regression model results in Table 4 shows that the F-statistic is low but the p-value for the overall model indicates that the model is statistically significant.

5.1.2 Fixed Effects Linear Regression Model

In order to account for commonalities among plans within a GCD over time, a fixed effects model was built. This model controls for correlation across plans belonging to the same GCD, as well as those submitted in the same year. As we can see from Table 5, there were no statistically significant coefficients in this model so we cannot reject the null hypothesis that the estimates are equal to zero.

Table 5: Fixed Effects Model Results accounting for GCD Name and Plan Year

Fixed Effects (GCD Name and Year) Results:				
Residuals:				
Min	**1Q**	**Median**	**3Q**	**Max**
-386850	-20773	0	25349	1000632
Coefficients	**Estimate**	**Std. Error**	**t Value**	**Pr(>|t|)**
District Count	-108.602	194.535	-0.558	0.577
Storage Count	341.164	926.562	0.368	0.713
Permit Count	-177.093	382.031	-0.464	0.644
Management Count	-50.895	358.762	-0.142	0.887
Aquifer Count	37.848	213.511	0.177	0.86
Recharge Count	278.822	623.809	0.447	0.655
Withdraw Count	-8.612	2190.713	-0.004	0.997
Shortage Count	-1585.378	3226.754	-0.491	0.624
Conservation Count	-145.755	488.276	-0.299	0.766
Resource Count	-1721.304	1653.511	-1.041	0.299
Groundwater Count	223.424	213.789	1.045	0.298
Precipitation Count	-896.365	1666.275	-0.538	0.591
Supply Count	960.506	1402.943	0.685	0.495
Demand Count	-499.761	873.807	-0.572	0.568
Words Per Page	-254.736	289.187	-0.881	0.38
Total Wordcount	1.679	3.893	0.431	0.667
Pages	-348.939	936.674	-0.373	0.71
Residual standard error: 120600 on 161 degrees of freedom				
Multiple R-squared (full model): 0.5289 Adjusted R-squared: 0.1368				
Multiple R-squared (projected model): 0.06489 Adjusted R-squared: -0.7134				
F-statistic (full model):1.349 on 134 and 161 DF, p-value: 0.03467				
F-statistic (projected model): 0.6572 on 17 and 161 DF, p-value: 0.8407				

The fixed effects model shows that there is some commonality among the plan documents for the same GCD and year that the earlier regression model attributes to the other predictor variables. This means that when the model accounts for commonalities within GCDs and within the same submission year, none of the predictors have a statistically significant association with the response variable. The outcome of this model and that of the linear regression model are good indicators that this is a classification problem rather than a regression problem. This is because ultimately, the goal is to understand whether the plan content can classify the GCDs into the categories of water supply shortage vs water supply surplus. It is not a regression problem because determining if the variables can predict the value of water supply need is less informative when it comes to analyzing the overall management plan content. In addition, the regression results thus far do not support a causal relationship between any of the explanatory variables from the GCD management plans and the water supply variable.

5.1.3 Logistic Regression Model Results

In the logistic regression model, the predictor variables that were statistically significant at $p \leq .001$ were the count of “Aquifer” and “Recharge”. The model also shows the count of “Groundwater” statistically significant at $p \leq .01$ and the count of “Precipitation” statistically significant at $p \leq .05$. In this model the dependent variable is shortage = 0 and surplus =1. Therefore, the variables that are negatively associated with water supply need (Aquifer Count, Precipitation Count) suggest a reduced probability of observing a surplus, and those with positive coefficient estimates (Recharge Count, Groundwater Count) suggest an increased probability of observing a surplus. Converting the coefficient estimates to marginal effects, the odds of a water supply surplus decrease by a factor of 0.98 (or 2.1%) if aquifer count increases by one and the odds of a water supply surplus increase by a factor of 1.1 (or 9.9%) if Recharge count increases by 1.

Table 6: Logistic Regression Model Results

Logistic Regression Results:					
Deviance Residuals:					
Min	**1Q**	**Median**	**3Q**	**Max**	
-2.4545	-0.6732	0.2077	0.6736	2.614	
Coefficients	**Estimate**	**Std. Error**	**t Value**	**Pr(>|t|)**	
(Intercept)	0.971000	1.610000	0.603000	0.546370	
District Count	-0.006206	0.003502	-1.772000	0.076400	.
Storage Count	0.005823	0.017200	0.339000	0.734940	
Permit Count	0.009127	0.008206	1.112000	0.266070	
Management Count	0.008752	0.008272	1.058000	0.290040	
Aquifer Count	-0.021540	0.004613	-4.669000	0.000003	***
Recharge Count	0.094170	0.021790	4.322000	0.000015	***
Withdraw Count	0.002426	0.039190	0.062000	0.950640	
Shortage Count	0.080760	0.188700	0.428000	0.668630	
Conservation Count	-0.008062	0.009930	-0.812000	0.416880	
Resource Count	0.018890	0.025080	0.753000	0.451460	
Groundwater Count	0.014840	0.005405	2.745000	0.006050	**
Precipitation Count	-0.069830	0.033550	-2.082000	0.037380	*
Supply Count	-0.021650	0.026130	-0.829000	0.407310	
Demand Count	-0.039920	0.023390	-1.707000	0.087860	.
Words Per Page	0.001455	0.005993	0.243000	0.808150	
Total Wordcount	0.000029	0.000088	0.328000	0.742870	
Pages	-0.026450	0.020900	-1.266000	0.205640	
Test Error Rate: 21.88%					
Null deviance: 408.99 on 295 degrees of freedom					
Residual deviance: 295.57 on 278 degrees of freedom					
AIC: 331.57					

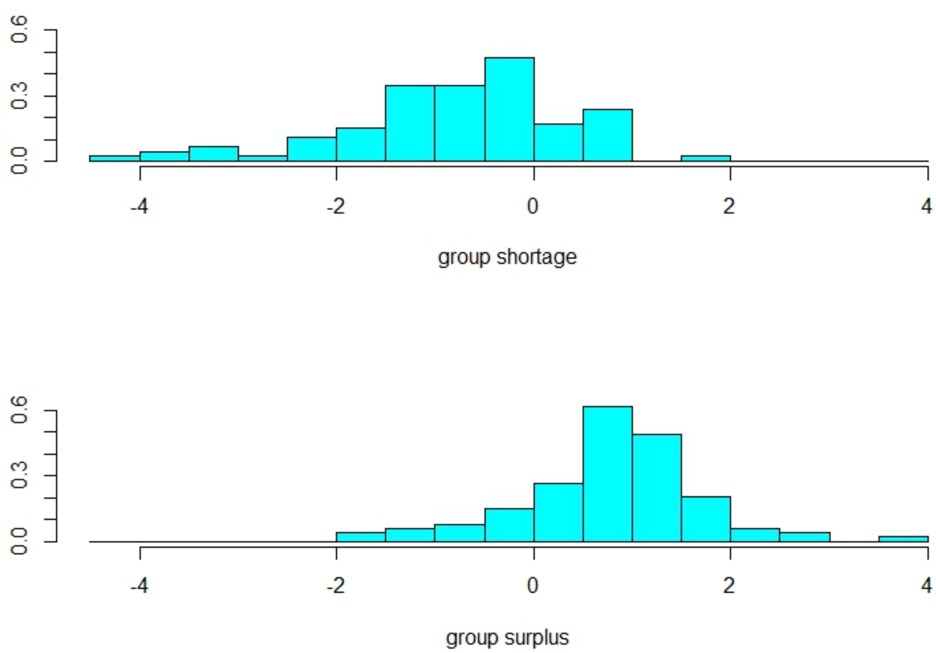
Null deviance is a measure of statistical significance for a model with one constant coefficient, because we have several coefficients in our model the residual deviance is more helpful. The chi-squared result for null deviance was 1.21×10^{-5} and the chi-squared result for residual deviance was 0.22. This means that the p-value for the model was 2.0×10^{-16} so we can reject the null hypothesis that the addition of the coefficients past the first constant coefficient add nothing to the explanation of the model. The logistic regression model results reported an overall test error rate of 21.88% which was lower than that of the LDA model that is covered in the next section.

Unlike linear regression, the logistic regression model cannot incorporate fixed effects to control comprehensively for time-invariant GCD characteristics, or for shocks over time that are common across GCDs. Thus, due to concerns about omitted variables bias, we cannot interpret these estimated effects as causal effects. Taken together, the regression results do not suggest any consistent relationship between water supply shortage/surplus and the independent variables (summarizing the content of the GCD management plans).

5.2 Linear Discriminant Analysis Model Results

In the LDA model the overall percentage of correct predictions in the test data was 75%, which leaves a test error of 25%. The model was correct 72.9% of the time when predicting a water supply surplus and the model was correct 78.4% of the time when predicting a water supply shortage. Figure 13 is a graphical representation of the correct prediction results of the test data. We can see from the figure that most of the incorrect predictions for shortage are not far off from the correct class, as nearly all of the incorrect observations for the shortage group, which is represented as negative numbers in this case, are less than 1. On the other hand, in the surplus group which is represented as positive here, none of the observations are smaller than -2 and most of the incorrect observations are clustered around the zero point, as we also saw with the shortage group observations. This is a great example of an instance where SVM models can help optimize the data to solve the classification problem because the incorrectly classified plans are clustered near to the class divide. This will be discussed in more detail in the SVM results section.

Figure 13: LDA model correct predictions plot



Interestingly, in Table 7 we can see that the two most influential coefficients belong to Precipitation count and Shortage count. Precipitation count is negative so it is predicting a shortage as we would assume from the other model results and prior examinations of the dataset. However, in the surplus group, shortage count emerged as a strong predictor of water supply surplus. This is interesting because it would indicate that management plans that emphasized shortage belong to GCDs that actually show a water supply surplus.

Table 7: LDA Model Results

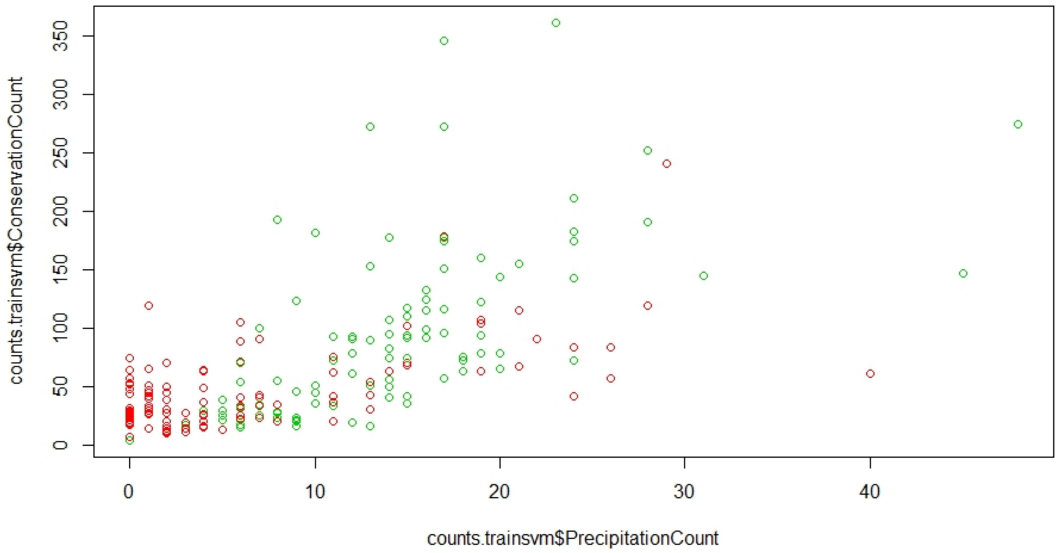
Prior probabilities of groups:	
shortage	**surplus**
0.465	**0.535**
Coefficients of linear discriminants:	
	LD1
District Count	-0.0045
Storage Count	0.0014
Permit Count	0.0028
Management Count	0.0061
Aquifer Count	-0.0115
Recharge Count	0.0443
Withdraw Count	0.0125
Shortage Count	0.0811
Conservation Count	-0.0031
Resource Count	0.0097
Groundwater Count	0.0055
Precipitation Count	-0.0447
Supply Count	-0.0083
Demand Count	-0.0198
Words Per Page	0.0044
Total Wordcount	0.0000
Pages	0.0003
Model Test Error Rate: 25%	

The prior probabilities of groups show the breakdown between classes that already exists in the training data. This means that 53.5% of the overall observations are assigned to the surplus class and 46.5% of the observations are assigned to the shortage class.

5.3 Support Vector Model Results

As discussed in Chapter 4, Support Vector Machine models are classification models that attempt to find the most optimal separation between classes. To test the optimal class separation of the data I built three SVM models, one linear kernel model, one radial kernel model and one polynomial kernel model. Each model was first built with default estimates for cost, gamma and degree and then tuned to find the best values for the parameters. Before these models were constructed, based on the findings of the other models I assumed that the linear kernel model would return the most optimal results but in fact it was the radial kernel model that did the best job of classifying the data into the two classes. I believe that this is because the predictor variables that were chosen for this series of models (precipitation count and conservation count) consistently showed successful predictions of shortage in the other models, but the ability to predict a surplus was not accounted for. As we can see in Figure 14, it is difficult to image a linear decision boundary that would neatly separate the observations into the two classes.

Figure 14: Plot of SVM model training data, showing classes of Water Supply Need by count of Precipitation and count of Conservation



The initial linear kernel model was set to cost=.01, the initial radial kernel model was set to cost=1, gamma=1, and the initial polynomial kernel model was set to cost=1, gamma=.5 and degree =2.

5.3.1 LINEAR KERNEL MODEL RESULTS

The first SVM model to be built was the Linear Kernel model. It can be called the most straightforward of the SVM models because it only involves on tuning parameter (cost). Before the model was tuned, the overall percentage of correct predictions was 70.83%, with 63.4% correct predictions of a shortage and 76.4% correct predictions of a surplus. After tuning the overall percentage of correct predictions rose to 79.2%, with 87.8% correct when predicting a shortage but dropped to 72.7% correct predictions for surplus. These results are the difference between running the model with a cost parameter of .1 and the same model with a cost parameter of 1, which is an excellent demonstration of the power of the tuning parameter. In this scenario it means that the model with the lowest error does not predict the classification of surplus as well as shortage but overall the correct prediction rates are high for both classes combined.

Table 8: SVM Linear Kernel Full Tuned Model Results

- Sampling Method: 10-fold cross validation		
- Best Parameters: Cost = 1		
- Best Performance: 0.245		
- Test Error Rate: 20.8%		
- Detailed performance results:		
cost	**error**	**dispersion**
0.0001	0.525	0.19329023
0.0010	0.455	0.17865236
0.0100	0.265	0.13550318
1.0000	0.245	0.08644202
10.0000	0.27	0.11105554
100.0000	0.29	0.12649111
1000.0000	0.28	0.11105554

Table 8 displays the cost, error and dispersion results for the tuned SVM linear kernel model. The tuning process selected the cost of 1 as the optimal model, as mentioned above. The number of support vectors in the original model was 54, the number of support vectors in the tuned model was 99, this is in line with the original cost of only .01 and the tuned cost of 1 because the large the cost, the wider the maximal margin width and therefore the larger the number of support vector observations. The tradeoff for a higher percentage of correct classifications in the tuned model is lower model variance but higher model bias.

5.3.2 RADIAL KERNEL MODEL RESULTS

The “best” tuned model for radial kernel had a cost of 10 and a gamma of 0.1, and error of 0.19. The tuned model returned an overall percentage of correct predictions of 70.8%, with 65.5% correct when predicting a shortage and 78.5% correct predictions for surplus. Recall that the radial kernel is a highly localized model in the sense that it is only the nearest training observations that impact the class label of a test observation. This is ideal for our dataset because there are several observations that are not grouped according to a linear decision boundary, as we have observed from the LDA results and Linear Kernel results.

Table 9: SVM Radial Kernel Model Results

- Sampling Method: 10-fold cross validation			
- Best Parameters: Cost =10, Gamma = 0.1			
- Best Performance: 0.19			
- Test Error Rate: 29.2%			
- Detailed Performance Results:			
cost	**gamma**	**error**	**dispersion**
0.0001	0.01	0.515	0.14151953
0.0010	0.01	0.515	0.14151953
0.0100	0.01	0.515	0.14151953
1.0000	0.01	0.235	0.08834906
10.0000	0.01	0.195	0.04972145
100.0000	0.01	0.245	0.08644202
1000.0000	0.01	0.24	0.10488088
0.0001	0.1	0.525	0.12304019
0.0010	0.1	0.525	0.12304019
0.0100	0.1	0.525	0.12304019
1.0000	0.1	0.205	0.05502525
10.0000	0.1	0.19	0.06146363
100.0000	0.1	0.19	0.06146363
1000.0000	0.1	0.19	0.06146363
0.0001	0.5	0.535	0.11067972
0.0010	0.5	0.535	0.11067972
0.0100	0.5	0.535	0.11067972
1.0000	0.5	0.215	0.08181958
10.0000	0.5	0.215	0.08181958
100.0000	0.5	0.215	0.08181958
1000.0000	0.5	0.215	0.08181958
0.0001	1	0.53	0.11595018
0.0010	1	0.53	0.11595018
0.0100	1	0.53	0.11595018
1.0000	1	0.29	0.12649111
10.0000	1	0.265	0.09143911
100.0000	1	0.265	0.09143911
1000.0000	1	0.265	0.09143911
0.0001	2	0.545	0.10658851

Table 9: SVM Radial Kernel Model Results Continued

cost	gamma	error	dispersion
0.0010	2	0.545	0.10658851
0.0100	2	0.545	0.10658851
1.0000	2	0.5	0.12909944
10.0000	2	0.485	0.12483322
100.0000	2	0.485	0.12483322
1000.0000	2	0.485	0.12483322
0.0001	3	0.545	0.10658851
0.0010	3	0.545	0.10658851
0.0100	3	0.545	0.10658851
1.0000	3	0.53	0.10852547
10.0000	3	0.52	0.10593499
100.0000	3	0.52	0.10593499
1000.0000	3	0.52	0.10593499
0.0001	4	0.545	0.10658851
0.0010	4	0.545	0.10658851
0.0100	4	0.545	0.10658851
1.0000	4	0.54	0.10749677
10.0000	4	0.54	0.10749677
100.0000	4	0.54	0.10749677
1000.0000	4	0.54	0.10749677
0.0010	4	0.48	0.12064641
0.0100	4	0.48	0.12064641
0.1000	4	0.26	0.09660918
1.0000	4	0.25	0.08498366
10.0000	4	0.3	0.11055416
100.0000	4	0.31	0.11005049
1000.0000	4	0.3	0.11547005

Figure 15: Plot of SVM Radial Kernel Model Results Before Tuning (Precipitation Count and Conservation Count only)

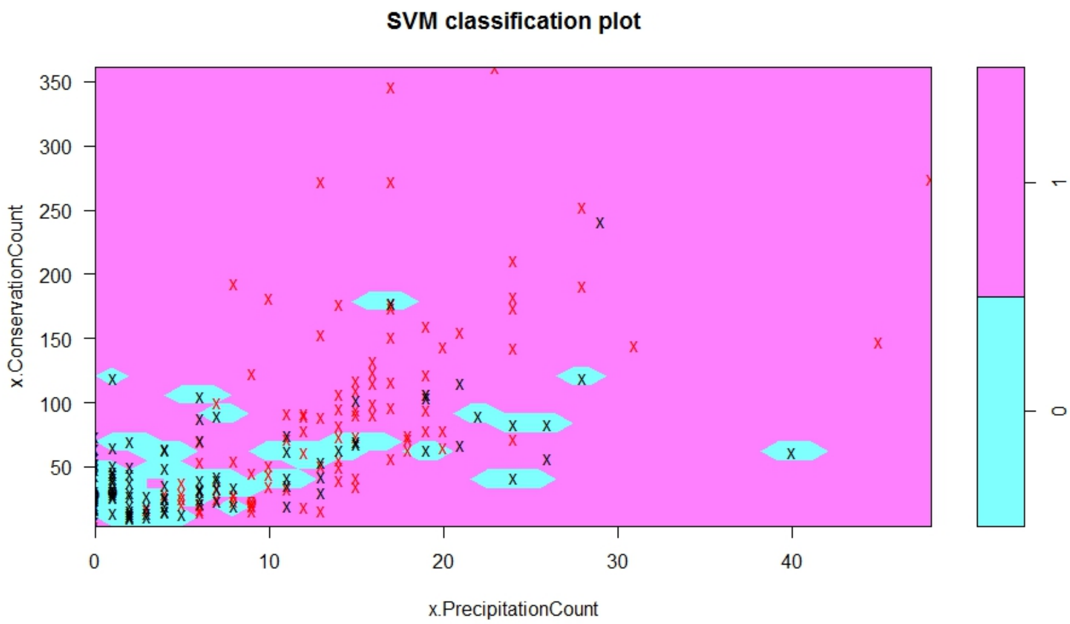
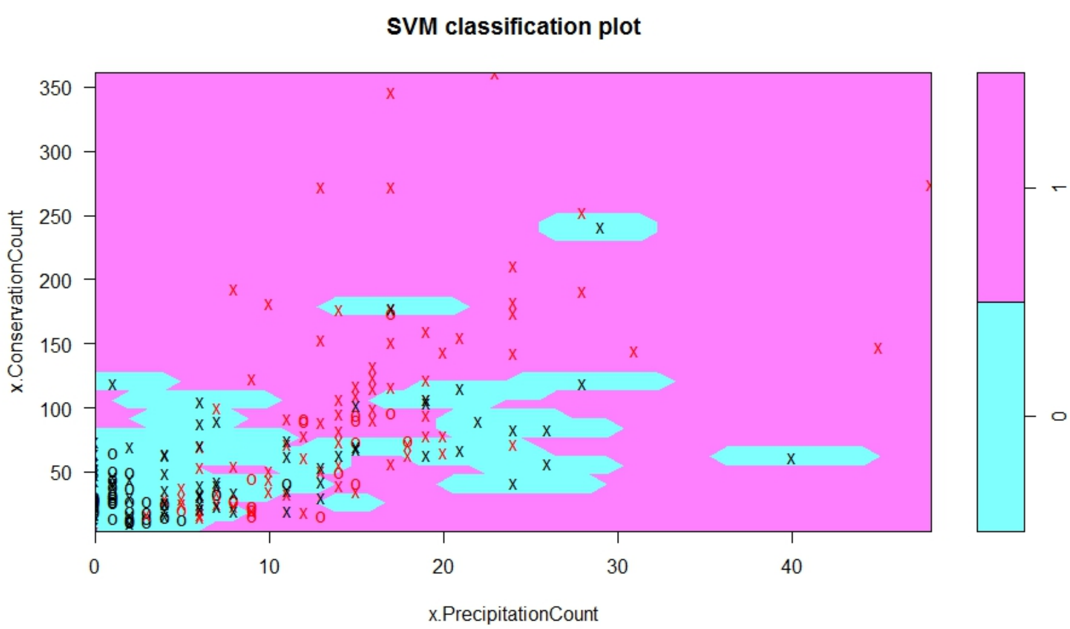


Figure 16: Plot of Tuned SVM Radial Kernel Model Results (Precipitation Count and Conservation Count only)



Before tuning, the radial kernel model had an overall correct prediction percentage of 64%. Figure 15 and Figure 16 provide an excellent example the impact of tuning parameters has on model fit. Figure 15 shows the radial kernel model plot for Precipitation Count and Conservation Count before the model has been tuned and Figure 16 shows the same results after the parameter tuning. It is obvious how much more accurately the model is able to separate the data into the two classes. In these plots the class of shortage is equal to one and the class of surplus is represented as zero.

5.3.3 Polynomial Kernel Model Results

The best tuned model for polynomial kernel had a cost of 10 and a gamma of 0.1, and error of 0.195, regardless of the degree setting. The tuned model returned an overall percentage of correct predictions of 70.8%, with 65.5% correct when predicting a surplus and 78.05% correct predictions for shortage. These class results appear to be almost exactly the inverse of the radial kernel model. That tells us that the surplus class observations are more difficult to pick out of the dataset when the decision boundary is linear. On the other hand, the shortage class is easier to separate according to a linear decision boundary which concurs with the linear kernel results.

Table 10: Polynomial Kernel Model Results

- Sampling Method: 10-fold cross validation				
- Best Parameters: Cost=10, Gamma=0.1, Degree = 1,2,3				
- Best Performance: 0.195				
- Test Error Rate: 29.2%				
- Detailed Performance Results:				
cost	**gamma**	**degree**	**error**	**dispersion**
0.1	0.1	1	0.255	0.10394977
1	0.1	1	0.21	0.05163978
10	0.1	1	0.195	0.05986095
0.1	0.5	1	0.495	0.08959787
1	0.5	1	0.21	0.06146363
10	0.5	1	0.215	0.06258328
0.1	1	1	0.51	0.06992059
1	1	1	0.27	0.11595018
10	1	1	0.255	0.08644202
0.1	0.1	2	0.255	0.10394977
1	0.1	2	0.21	0.05163978
10	0.1	2	0.195	0.05986095
0.1	0.5	2	0.495	0.08959787
1	0.5	2	0.21	0.06146363
10	0.5	2	0.215	0.06258328
0.1	1	2	0.51	0.06992059
1	1	2	0.27	0.11595018
10	1	2	0.255	0.08644202
0.1	0.1	3	0.255	0.10394977
1	0.1	3	0.21	0.05163978
10	0.1	3	0.195	0.05986095
0.1	0.5	3	0.495	0.08959787
1	0.5	3	0.21	0.06146363
10	0.5	3	0.215	0.06258328
0.1	1	3	0.51	0.06992059
1	1	3	0.27	0.11595018
10	1	3	0.255	0.08644202

It is interesting that the “best model” (lowest error) polynomial kernel results are the same regardless of degree setting which means that additional polynomials not only do not improve the results but do not change them at all. As we can see from Figure 17 and Figure 18, shortage classifications are more amenable to a linear decision boundary because their values are higher so they take up more of the model space.

Figure 17: Plot of SVM Polynomial Kernel Model Results Before Tuning (Precipitation Count and Conservation Count only)

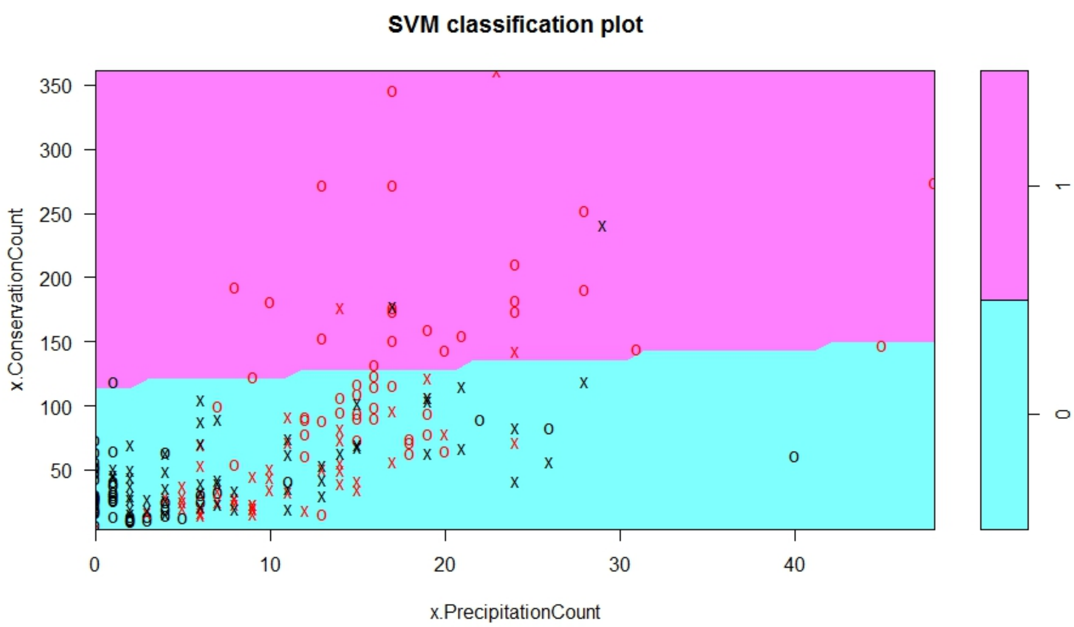
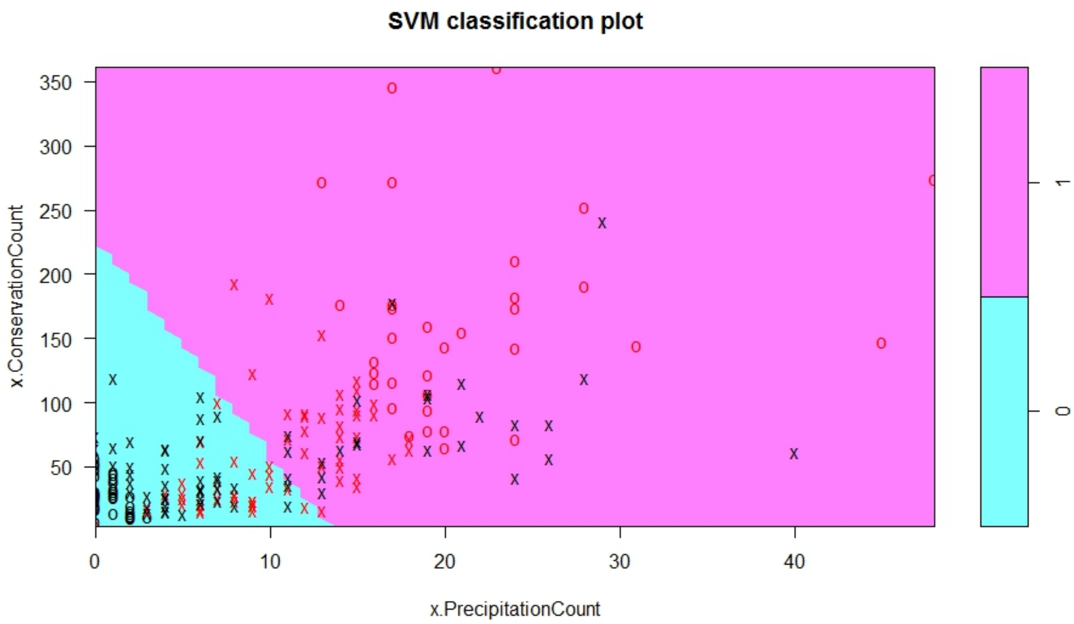


Figure 18: Plot of Tuned SVM Polynomial Kernel Model Results (Precipitation Count and Conservation Count only)



Even though the tuned polynomial model in Figure 18 is able to more accurately capture the class divide, there are still several misclassified observations.

CHAPTER 6: SUMMARY

6.1 Model Summary

To conclude, we first return to the original goals of this paper in Chapter 1.

1. Determine whether GCD management plans can be used to classify GCDs by water supply need in order to understand the impact of one on the other.
2. Discover what components have the strongest association with classification.

To achieve the first goal, the linear regression model results were promising at first but ultimately the fixed effects model results showed that once there were comprehensive, flexible controls for GCD characteristics and years, no predictor variables were significantly associated with the response variable. However, the original goal of this project is to predict the class of water supply need, not to establish causal relationships between predictors and water supply, or find the predictor variable(s) that best predict the status of water supply.

The classification models returned consistently high percentages of correct predictions. The model with the lowest test error rate was the linear kernel SVM model. This model had almost 90% correct predictions of observations belonging to the shortage class. Unfortunately, at the same time it only returned roughly 73% correct predictions for the surplus class. The radial kernel model returned the highest percentage of correct predictions for the surplus class (78.5%) but also had a high test error rate of almost 30% overall. Table 11 displays the high-level results for the SVM models.

Table 11: SVM Models Aggregated Results

Model Type	Overall Correct	Shortage Correct %	Surplus Correct %	Test Error Rate
Linear Kernel	79.20%	87.80%	72.70%	20.80%
Radial Kernel	70.80%	65.50%	78.50%	29.20%
Polynomial Kernel	70.80%	78.05%	65.50%	29.20%

The LDA model was also better at predicting a shortage (78.4% correct predictions) than a surplus (72.9% correct predictions), but the overall test error rate was higher (25%) than that of the linear kernel. In the logistic regression model results, the predictor variables that emerged as statistically significant at $P \leq .05$ or better were "Aquifer Count", "Precipitation Count", "District Count" and "Demand Count" for predicting the shortage (negative group) class. The predictor variables that emerged as statistically significant at $P \leq .05$ or better were "Recharge Count" and "Groundwater Count" for predicting the surplus (positive group) class.

An interesting discovery from the LDA model results in Table 7 is that the two most influential coefficients are Recharge and Shortage when predicting a surplus. This is interesting because it would indicate that management plans that emphasized shortage belong to GCDs that actually show a water supply surplus. This could indicate that GCDs that focus on planning for a water supply shortage are better prepared and therefore are able to maintain a water supply surplus. The same conclusion could be drawn about Recharge, in that the districts that focus planning efforts around recharge are actively avoiding a potential shortage situation. However, this result could also indicate that the heightened count of recharge is due to districts exploring artificial recharge projects, and in turn mitigating against potential drought conditions.

Overall, the first goal of this project was met but the second goal was not as obviously successful. The models were able to determine which predictor variables were the most significant within each model type but that alone does not provide a definitive variable that had the most impact. We can conclude that the predictors that are the strongest indicators of water supply class are as follows: Aquifer count, Recharge count, Precipitation count, Groundwater count, and Conservation count. However, the order of importance of these variables is debatable and it is also known that there is a common effect within GCDs and within the same plan year that mitigates the impacts of these variables.

If the TWDB or the GCD directors wish to perform this type of statistical learning analysis again in the future, there are a few items that could be improved in order to facilitate the overall process. Ideally, the response variable would be comprised of MAG values from GAM runs but there are two issues that prevent this. The first issue is that GAMs are a newer requirement, so the older plans do not contain that information. The second issue is that the GAM runs do not align with management plan submittal timelines so there are sometimes significant temporal discrepancies between the plan contents and the GAM results. As it is now, most districts MAG is projected for every 10 years for the next 5 decades into the future from the time of the GAM run. However, water supply need is projected for every 10 years from the date of the management plan for 5 decades into the future. For example, it creates a scenario where a 2016 GCD management plan with the most-recent GAM run in 2011 submits MAG data for the years 2010 through 2060 and Water Supply Need data for the years 2020 through 2070. If the GAM runs were aligned with the management plans it might make it easier to understand the resources of the district, in concert with the needs of the district.

In terms of data development for statistical modeling all of the newer plans are in great shape, but it would be recommended to reexamine the older documents. Management plans that are older than 2005 follow very different formats and do not include much of the same content, which makes them difficult to combine into a cohesive dataset. One of the stumbling points when assembling the dataset was access to the management plan PDFs. The plans were all publicly available from TWDB, which was excellent, and most were in great shape for data mining, however roughly 20-30% of the plans had significant issues in terms of accessibility. These issues ranged from password-protected PDFs to image-to-text extraction errors due to poorly scanned documents. The TWDB has largely corrected for this in the newer plans by releasing formatting guidance to assist the GCDs with plan development. An additional automated submission-validation tool would help to guarantee submission standards for digital uploads. Maximum/minimum page length requirements would help to standardize the dataset for statistical modeling but could be counterproductive for management because GCDs would be forced to cut off their plans at an arbitrary threshold. Lastly, water supply needs are projected into the future so it might be informative for GCDs to include the water supply need data tables from previous plans to explore the differences, if any, between past and current projections and to see if the predictions were accurate.

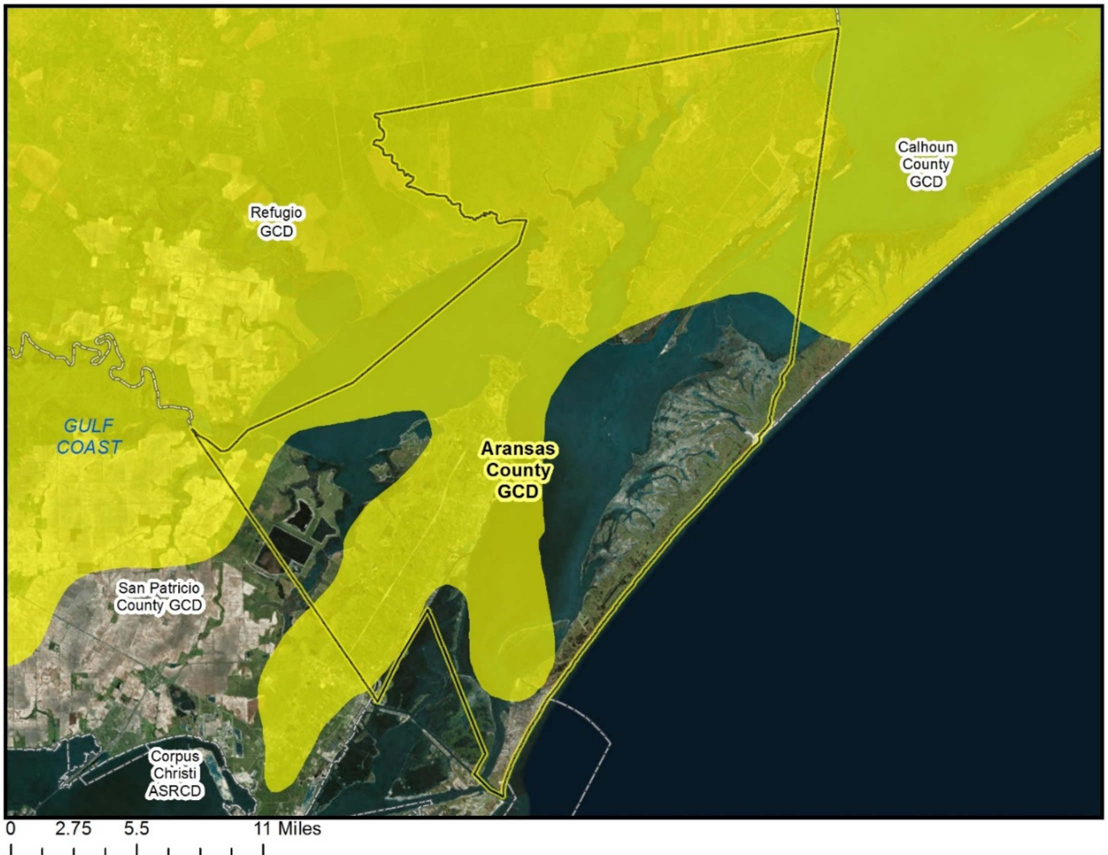
6.2 Overall Summary

This project set out to better understand the links between policy documentation and resource management in the field of groundwater. In some ways this was achieved and in some respects this project fell short of that goal. The model results were able to demonstrate that the overall predictor variables in the dataset can be used to successfully classify management plans into GCDs that report a water supply shortage or surplus. This means that there are clear links between enacted policy and the management of groundwater which does indicate that the management plans may be useful tools.

What we have learned through this project is that the management plans do have a strong connection to the water supply need classification of the resource, between 75% and 90% correct classifications depending on the model. We have also found that the most important predictors when determining class are shortage, recharge and groundwater when classifying as a surplus and precipitation, demand and aquifer when classifying a shortage. Now that we know that the management plans have a measurable impact on the classification of the GCD, and we have an idea of what factors are the most indicative of water supply need class, a general recommendation that we can make to the GCD directors would be to focus on recharge in their management plans. Another recommendation would be to emphasize planning for water supply shortages in order to either move towards, or maintain, a water supply need surplus. In conclusion, this project shows that the next set of proposed changes to GCD management should work within the current structure instead of the alternative. This is because we now know that GCD management plans are not just a statutory requirement, but they may be useful tools for resource management.

Appendix A

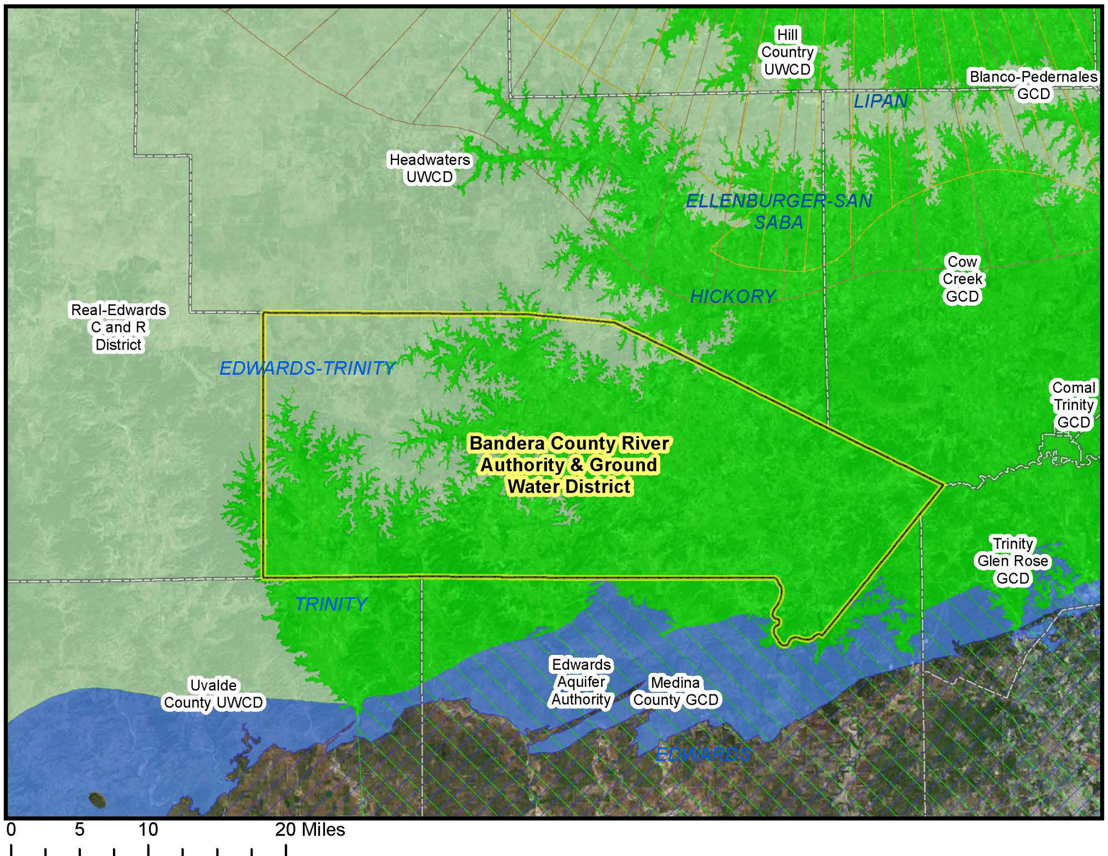
The following pages contain a map of each GCD accompanied by variable keyword pie charts, broken out by year where applicable. The source data for the GCD maps was provided by the TWDB. Please note the following GCDs have yet to submit their first GCD Management Plan: Aransas County GCD, Southwestern Travis County GCD.



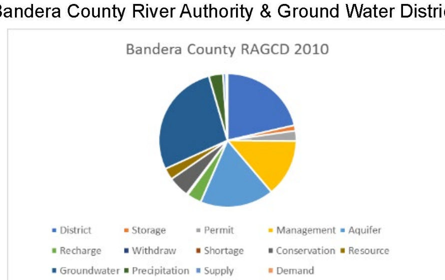
Aransas County GCD

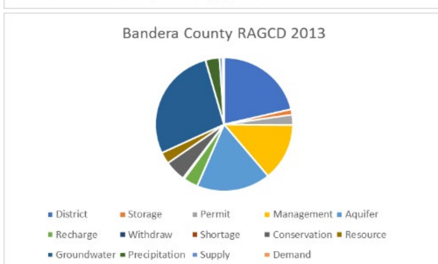
Aransas county GCD was created in 2016 and will submit its first GCD Management Plan in the near future.

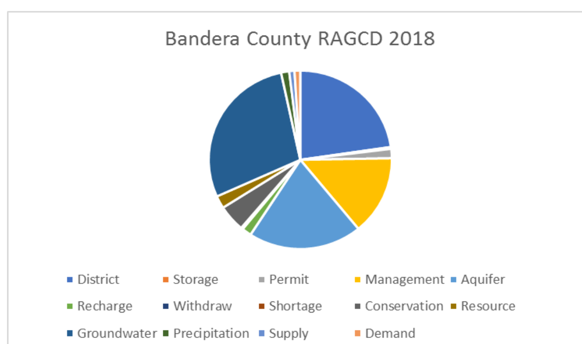
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



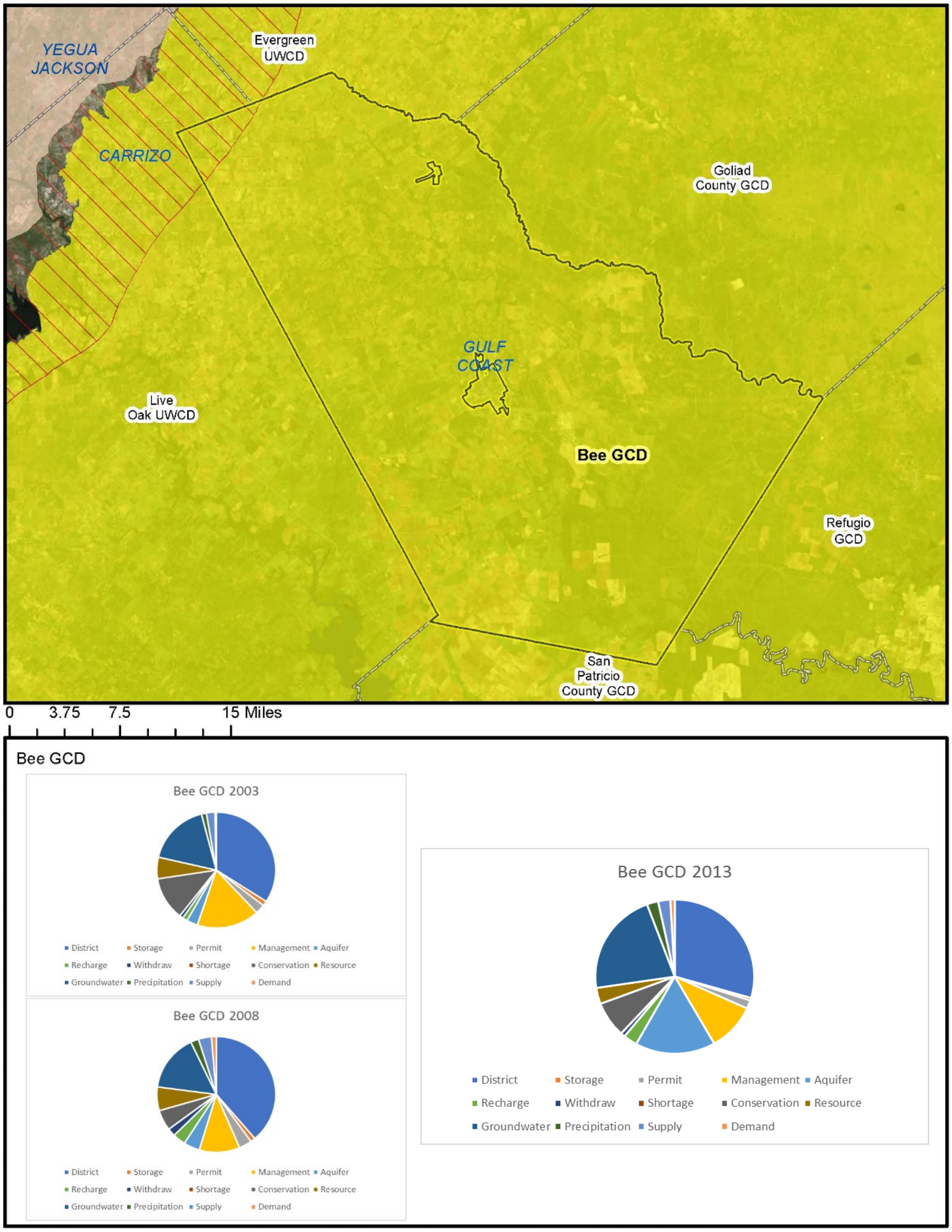
Bandera County River Authority & Ground Water District



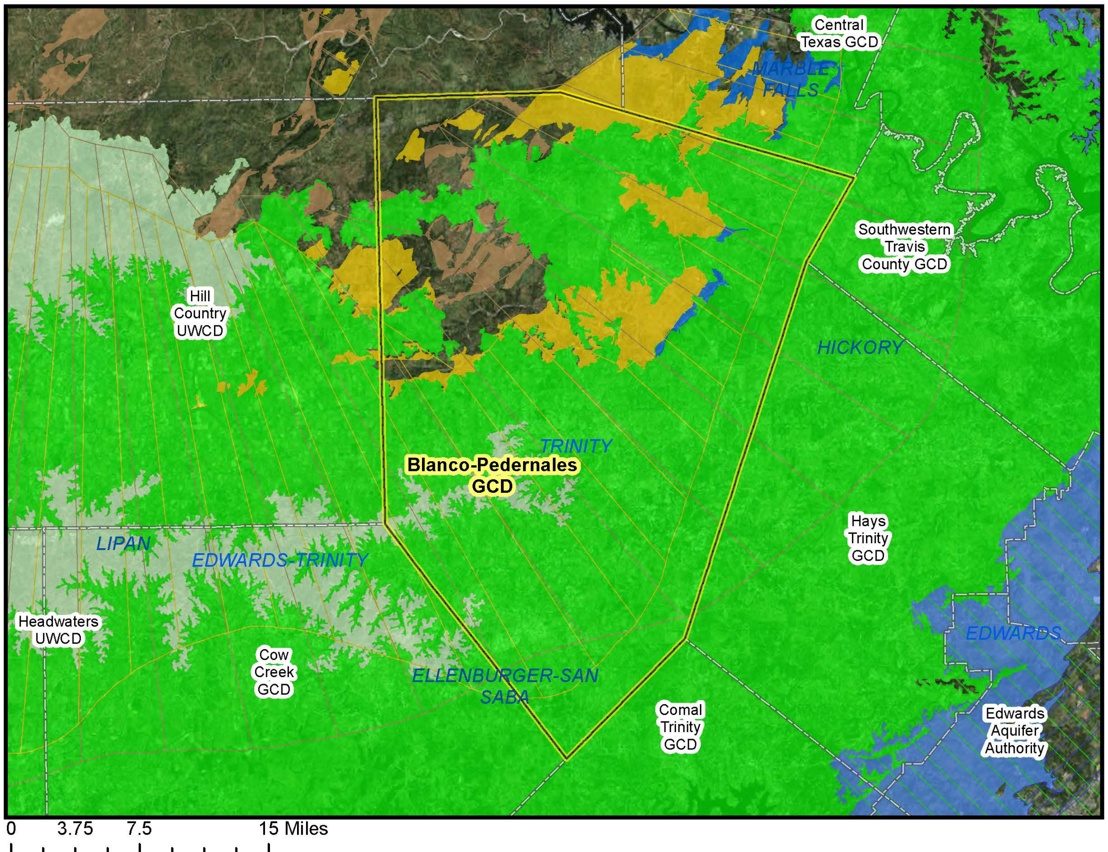




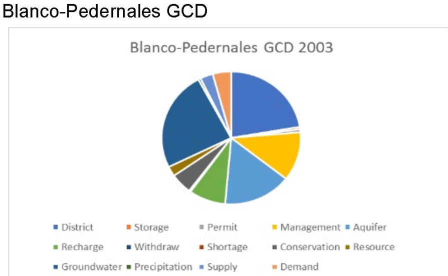
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

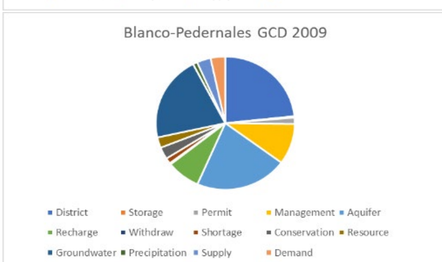


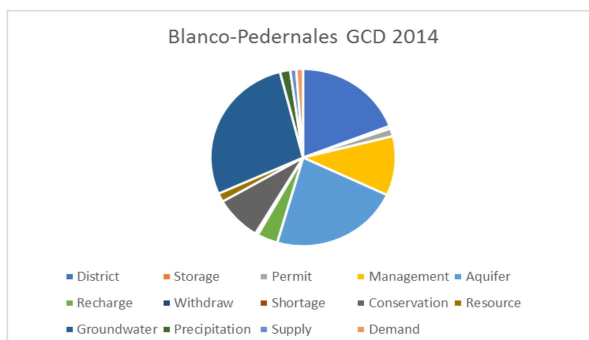
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



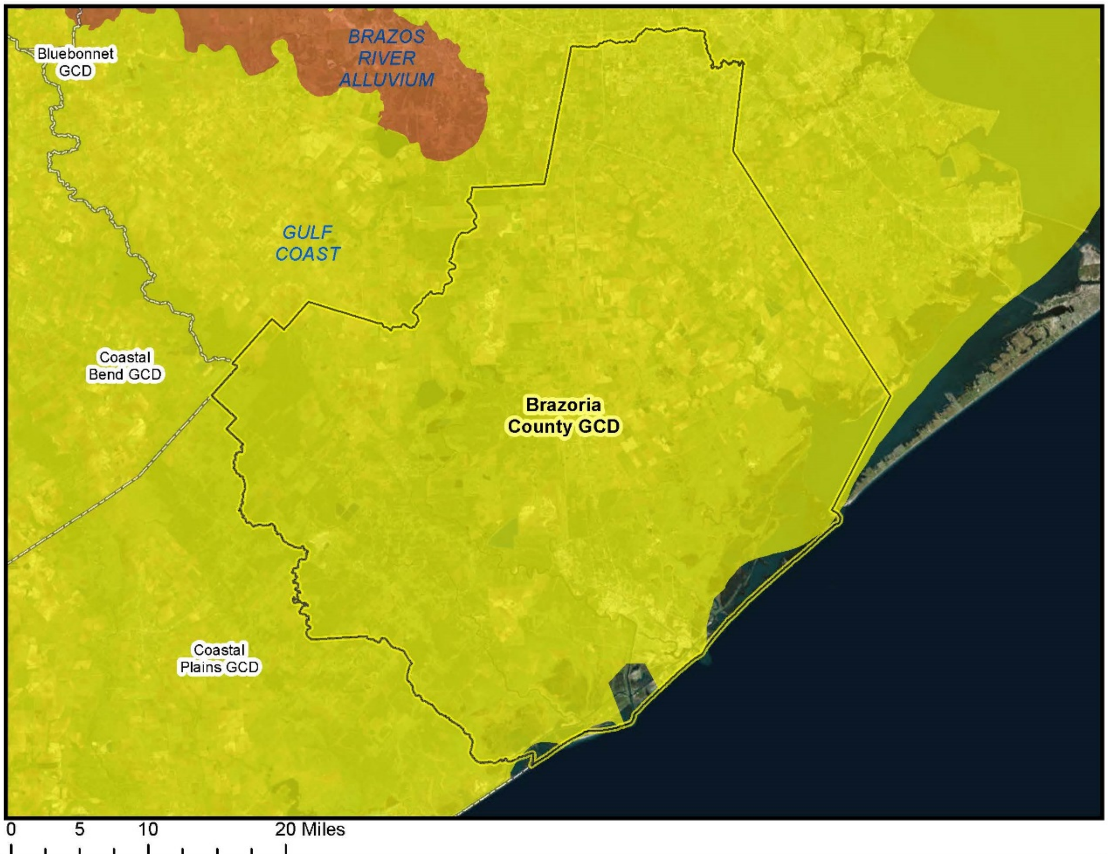
Blanco-Pedernales GCD



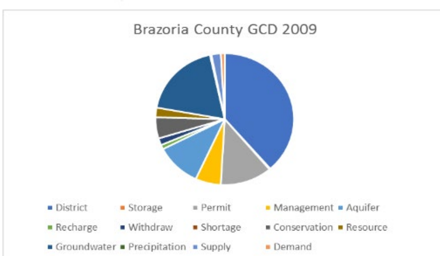


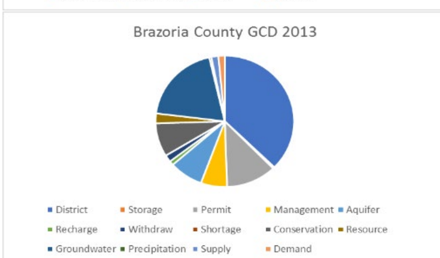


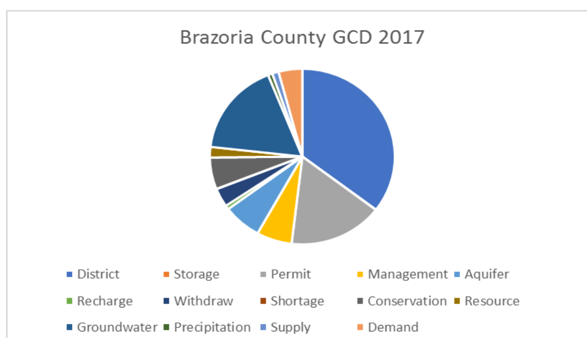
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



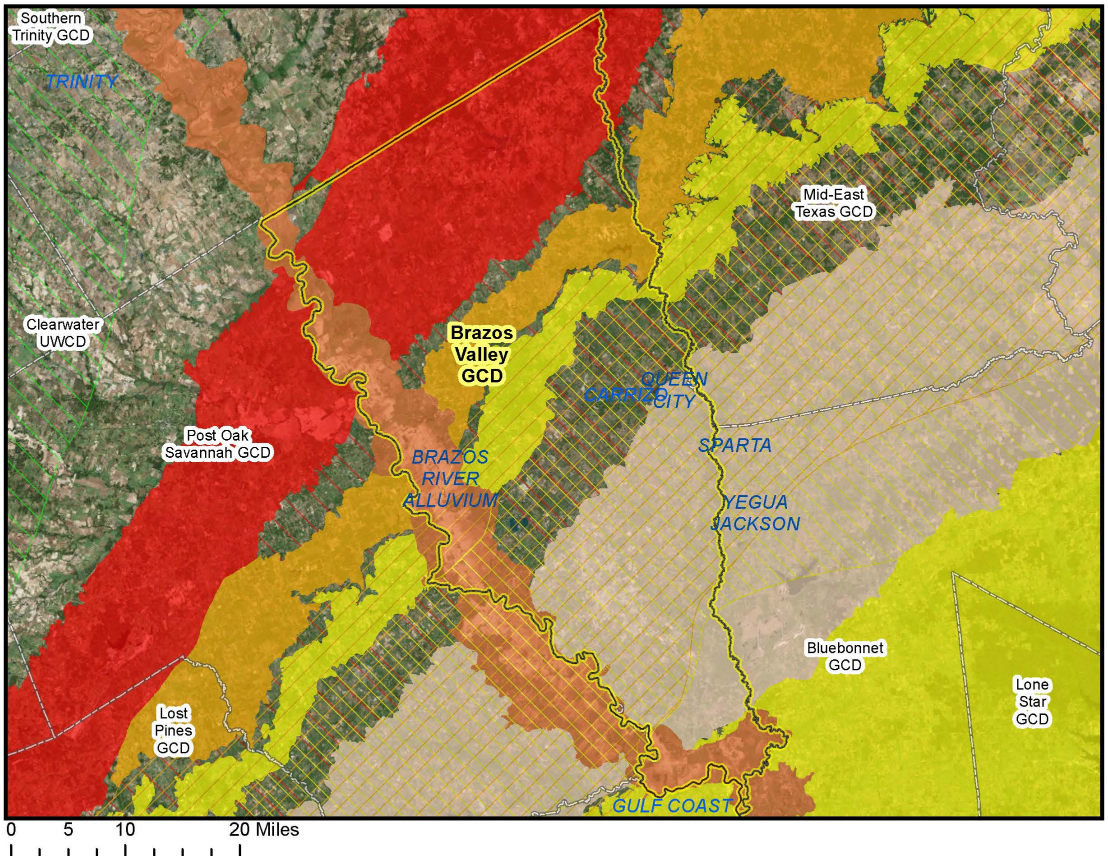
Brazoria County GCD



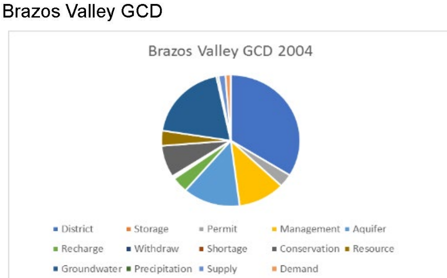


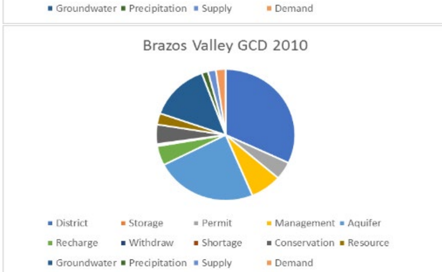


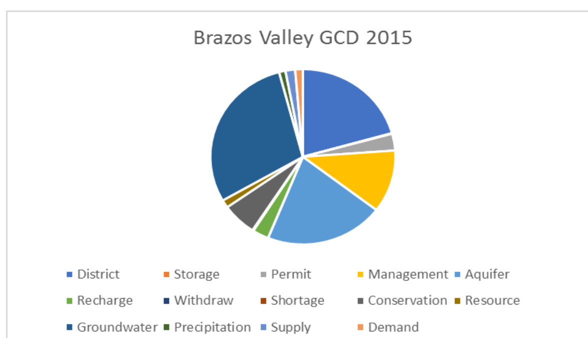
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



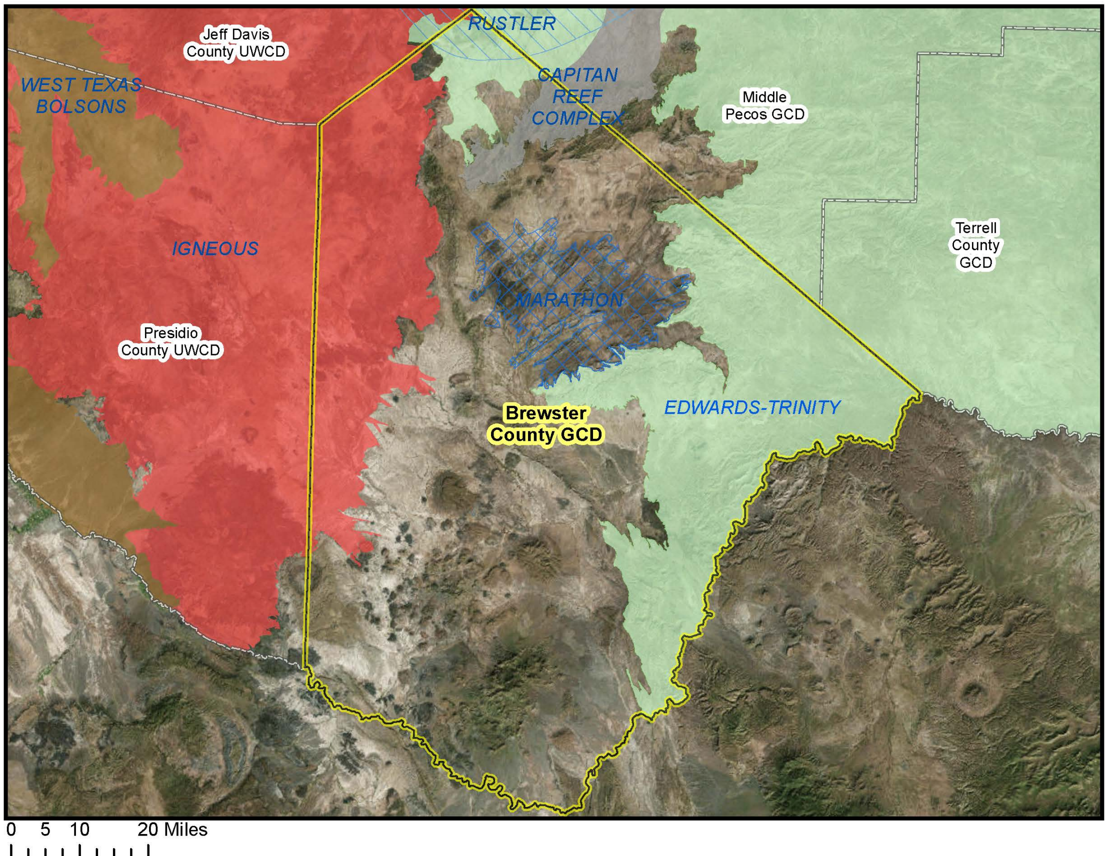
Brazos Valley GCD



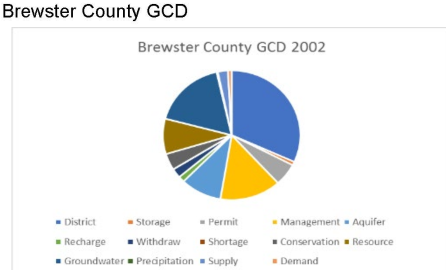


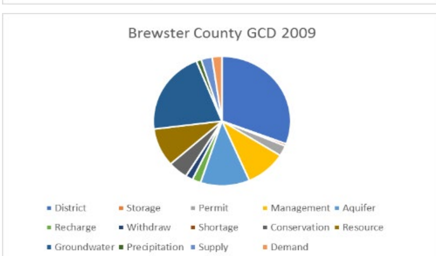


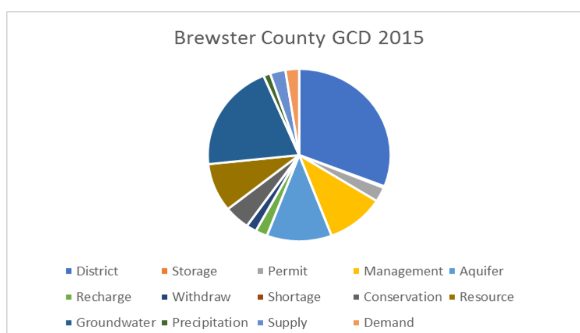
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



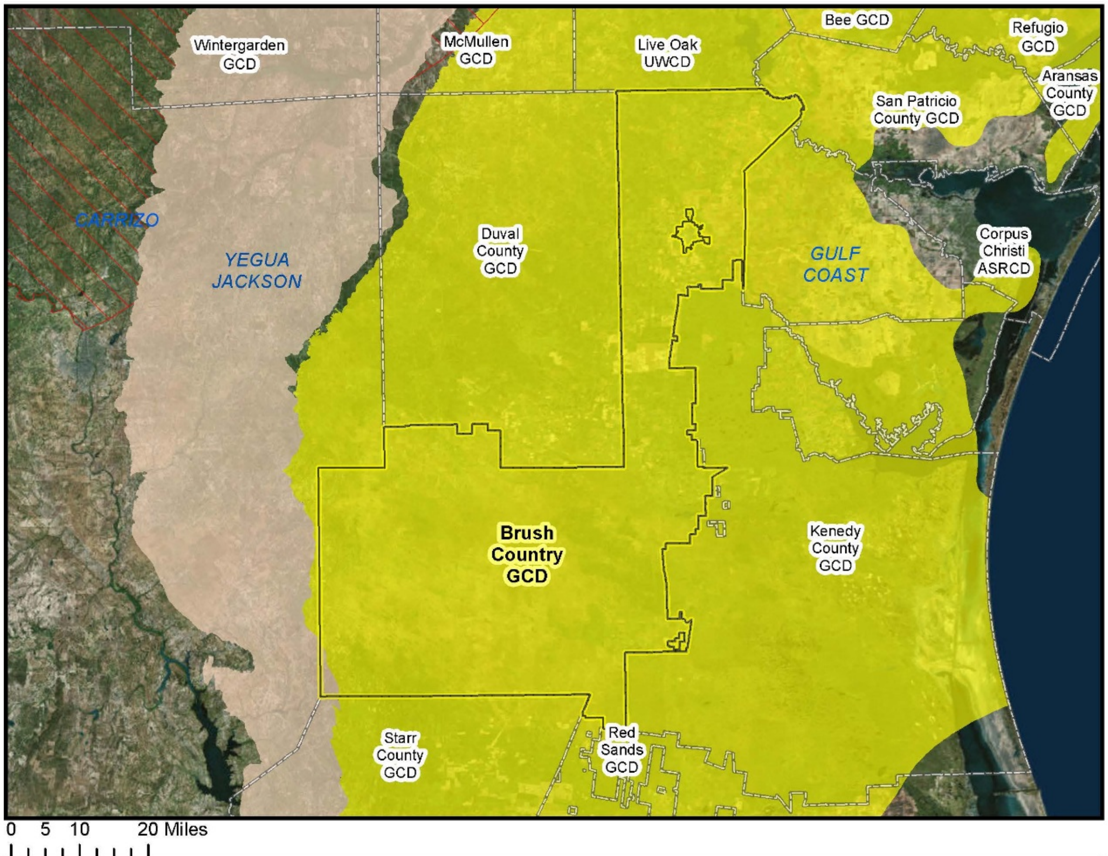
Brewster County GCD



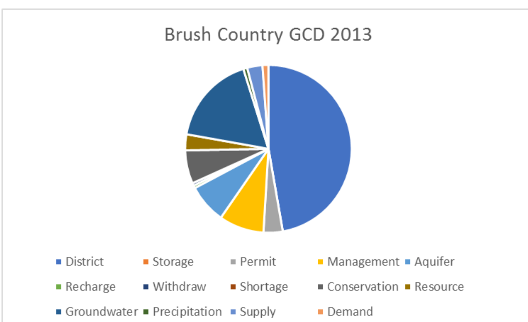


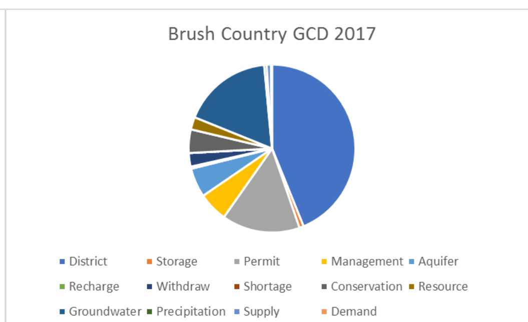


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

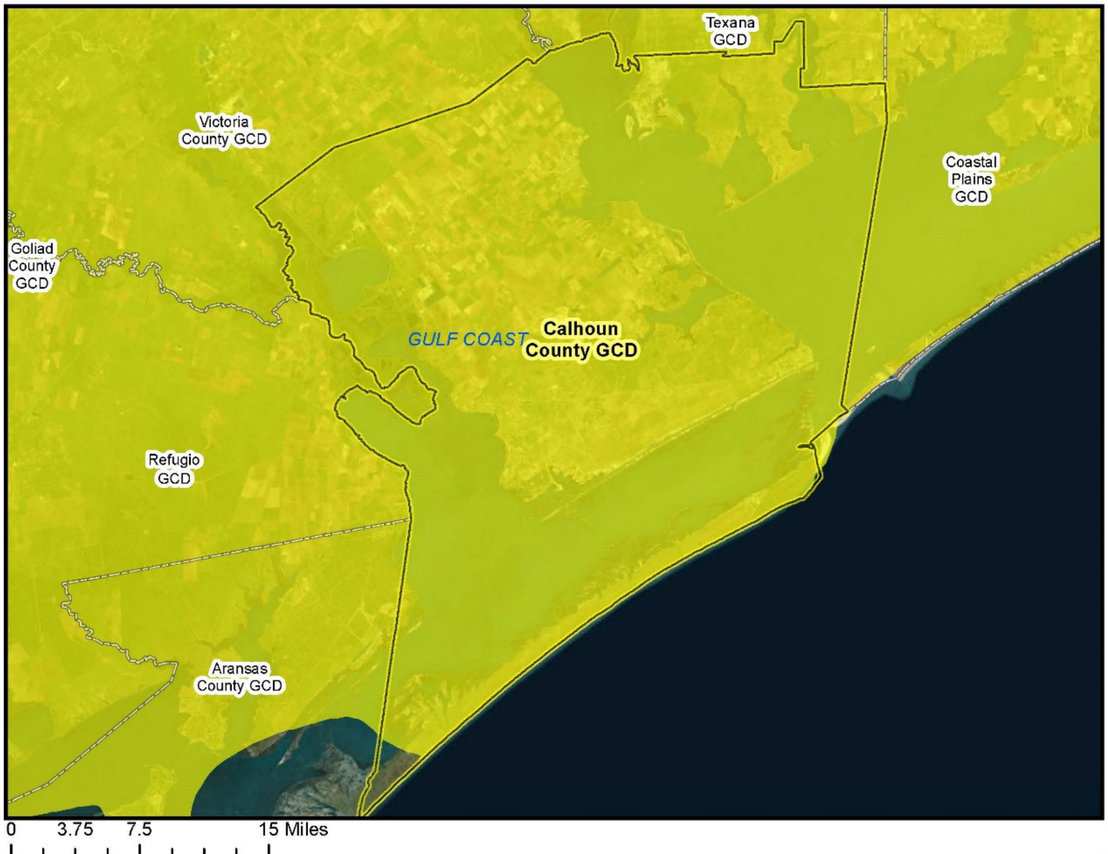


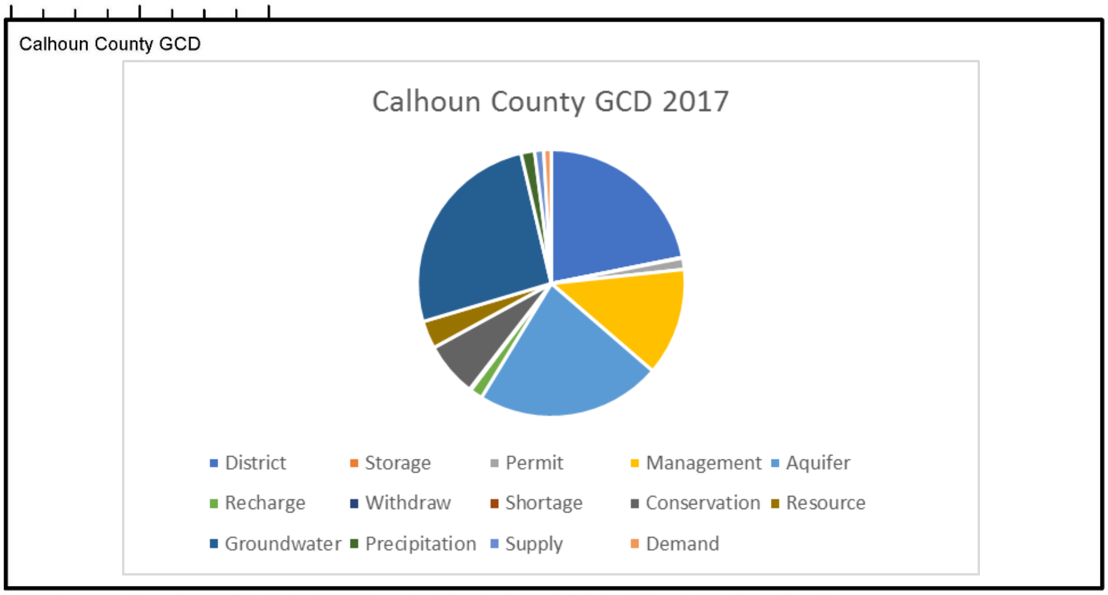
Brush Country GCD



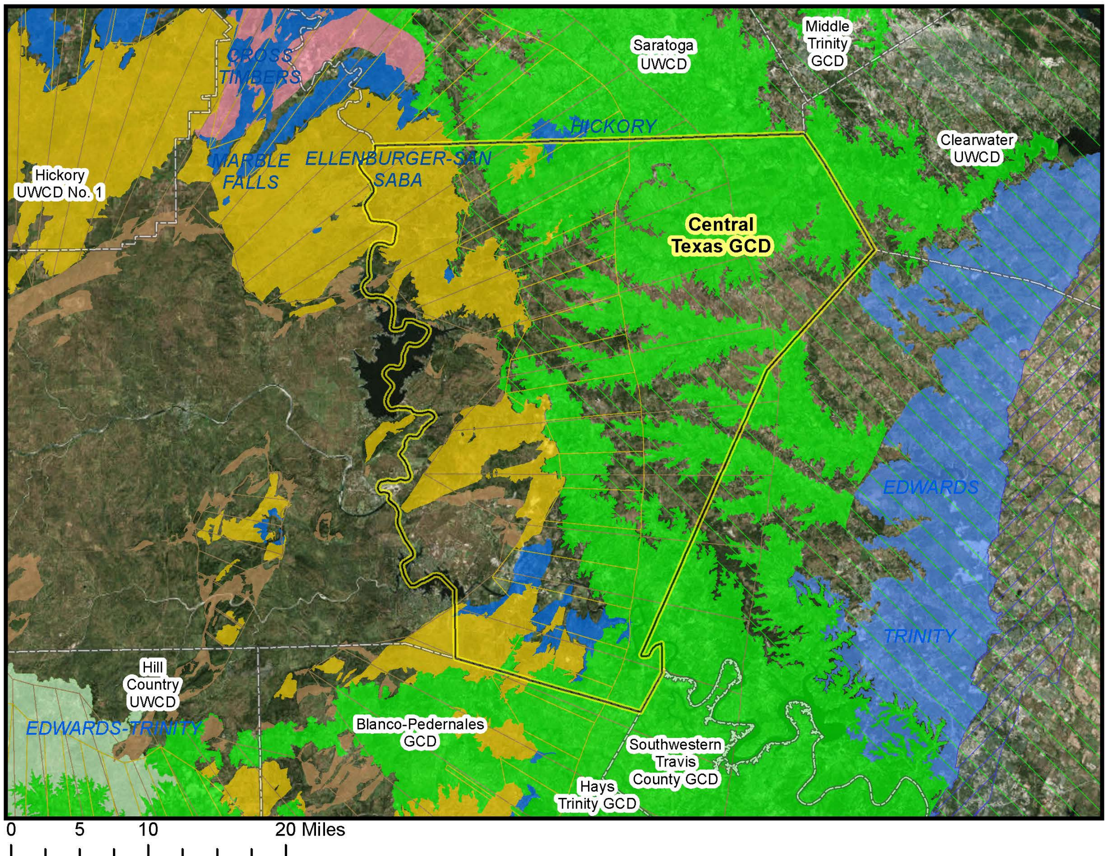


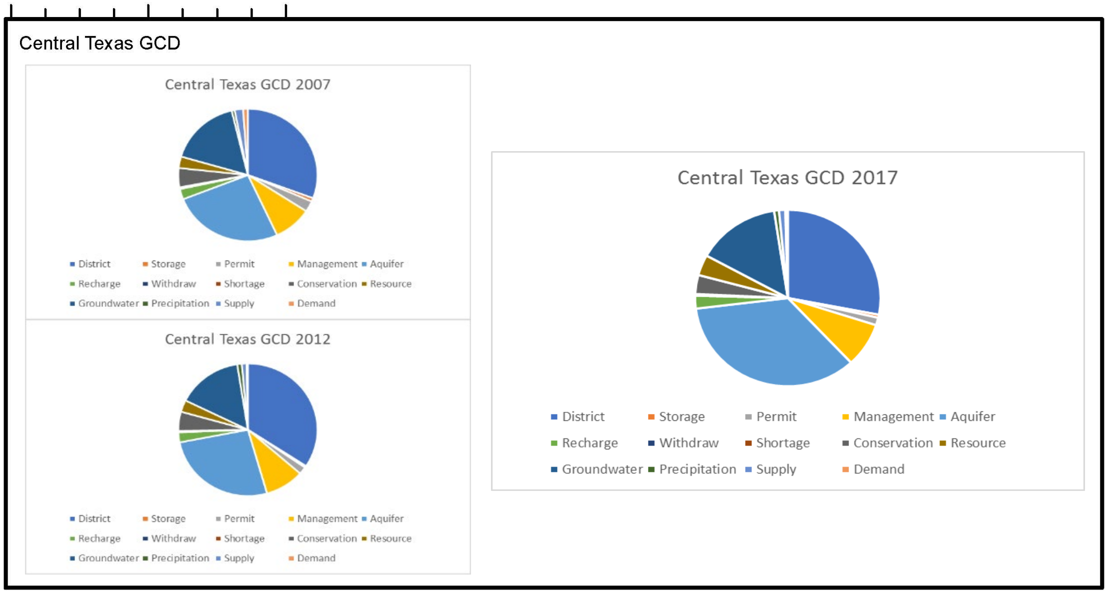
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



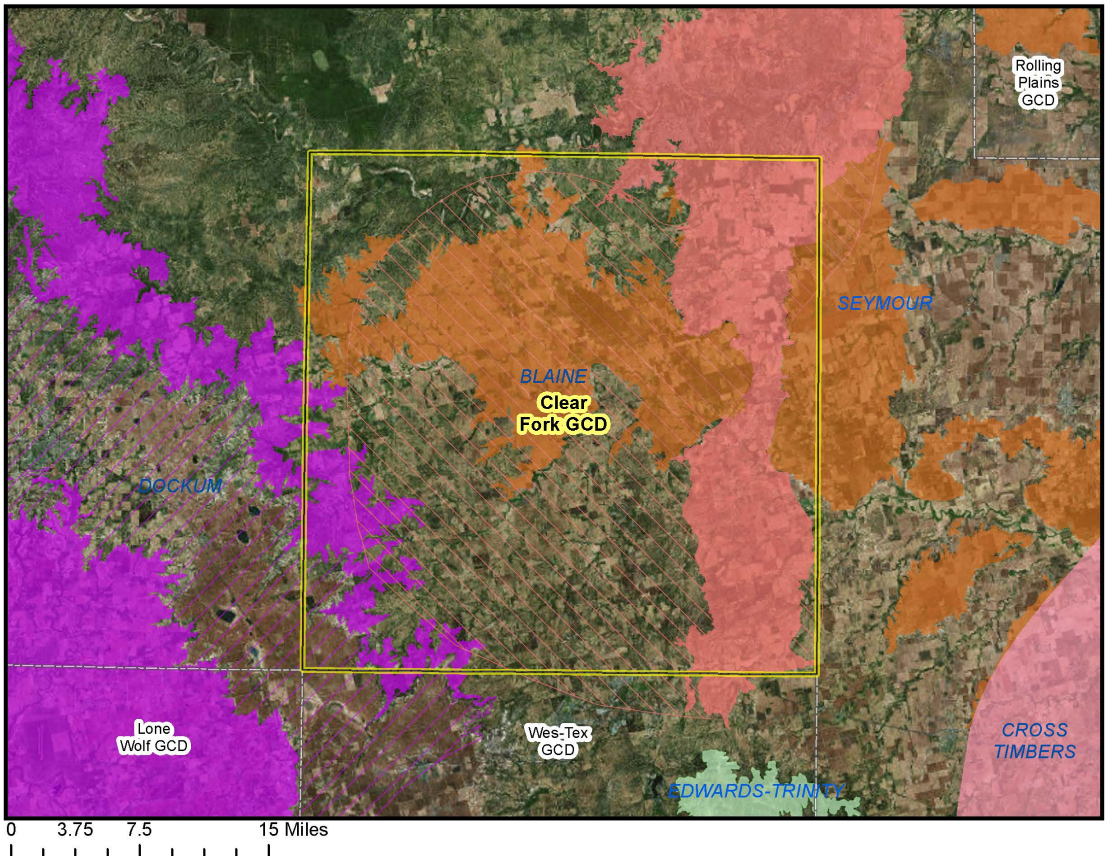


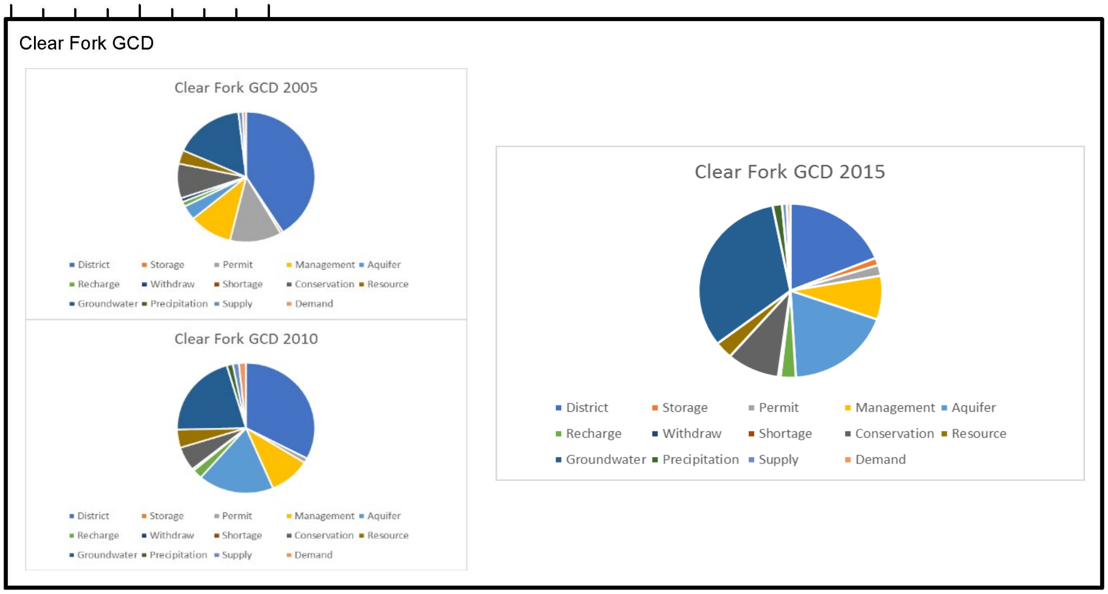
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



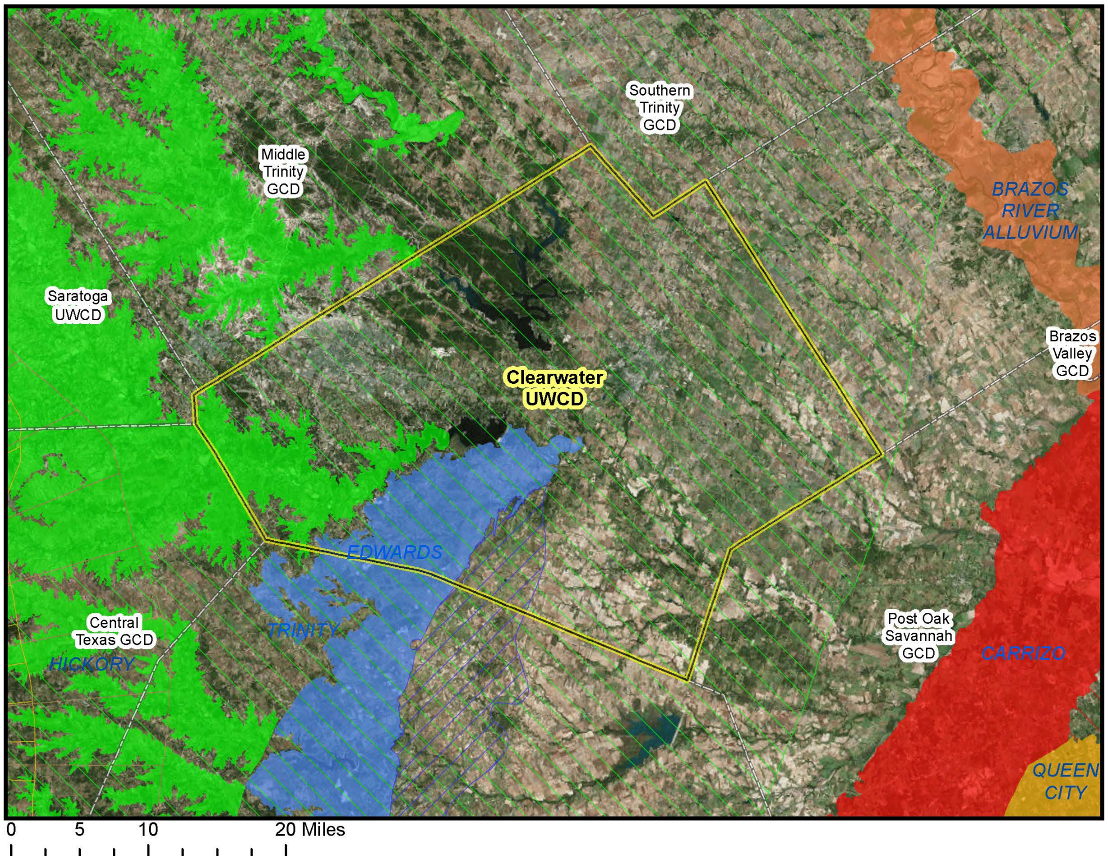


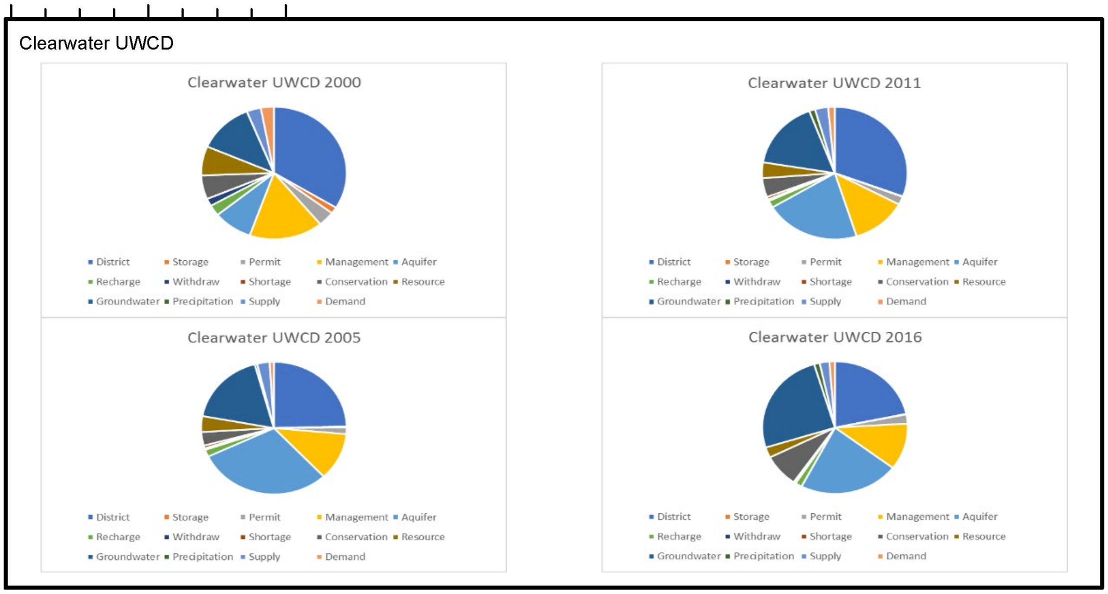
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



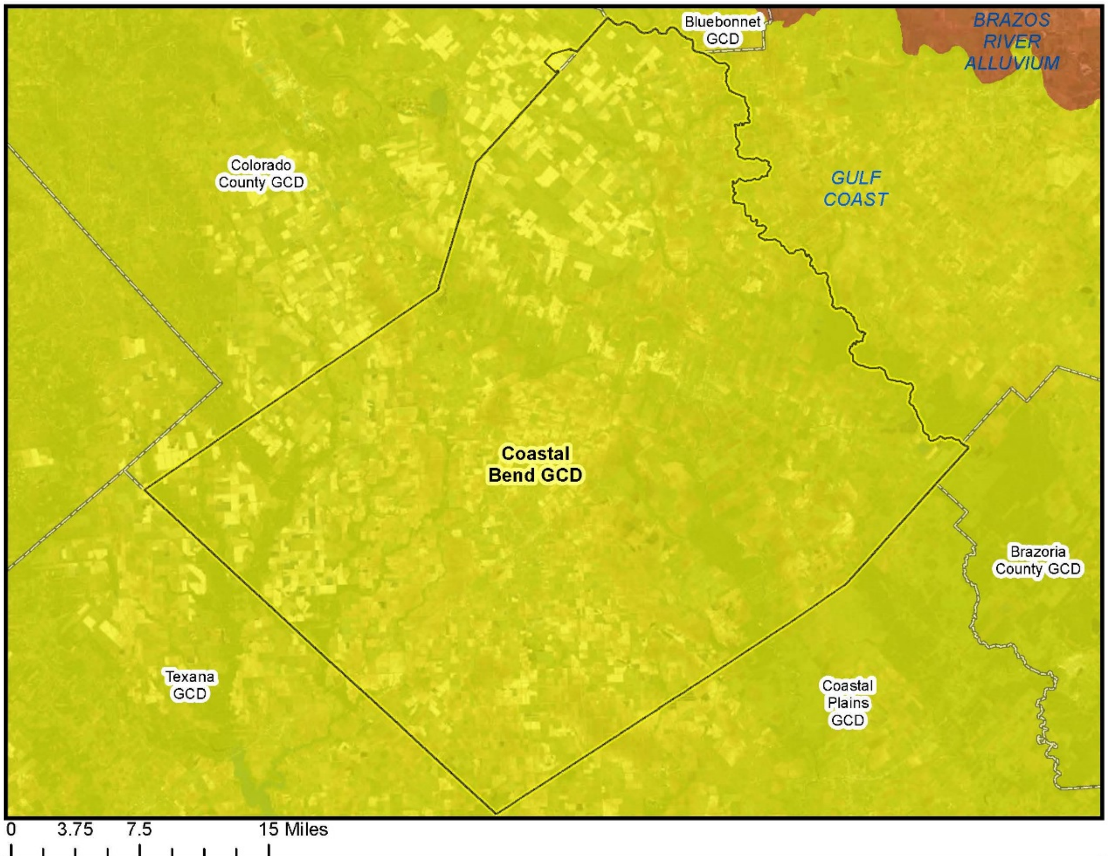


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

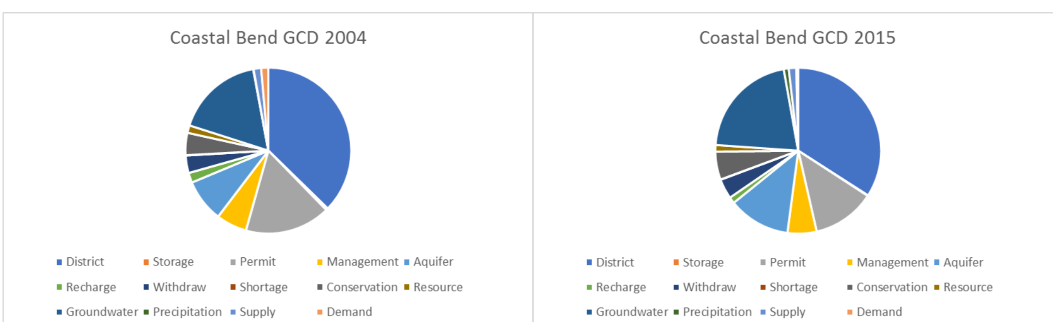




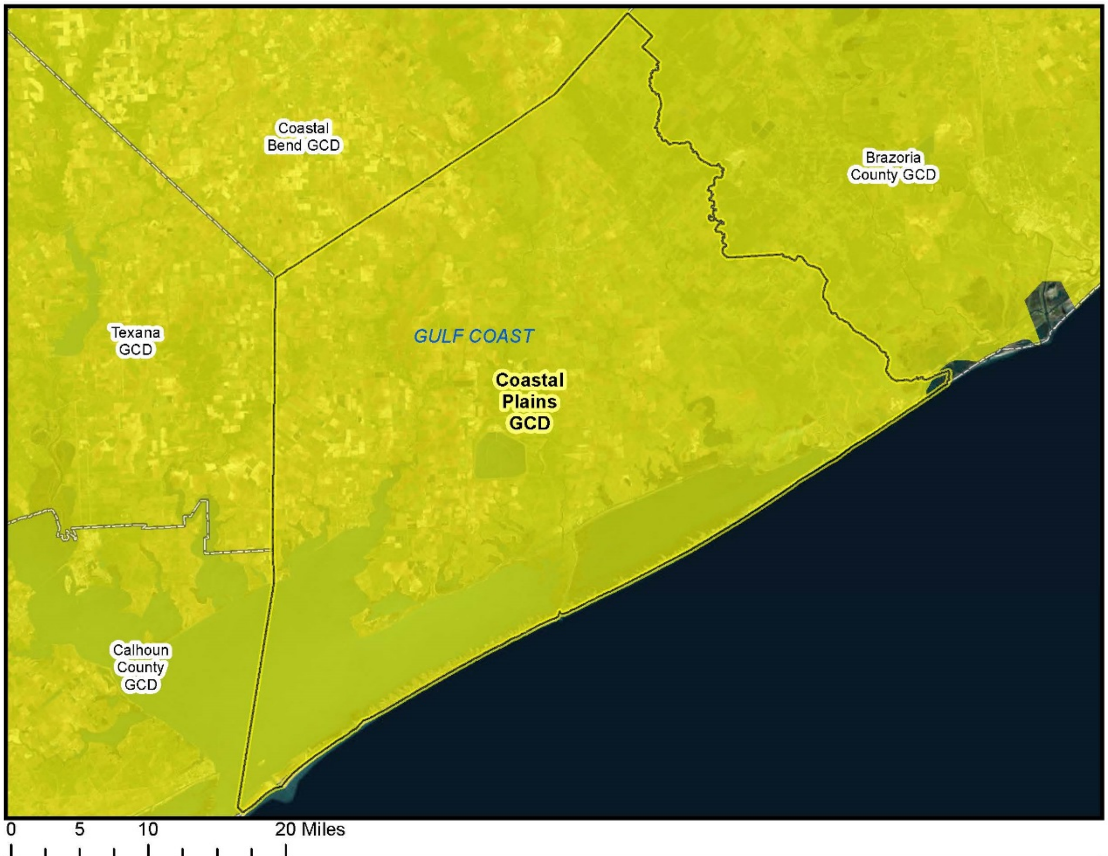
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

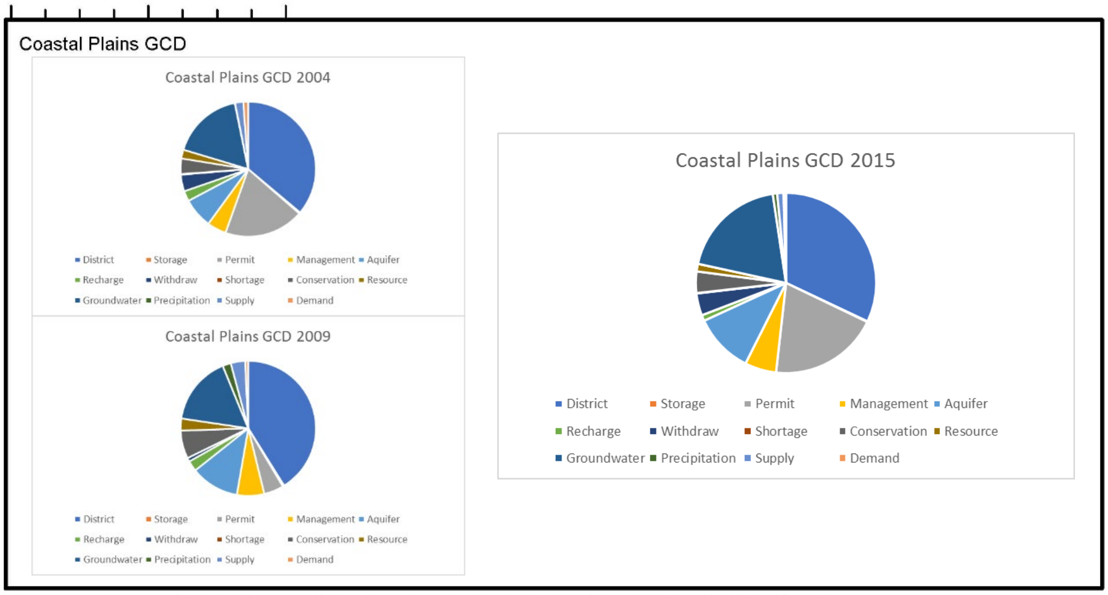


Coastal Bend GCD

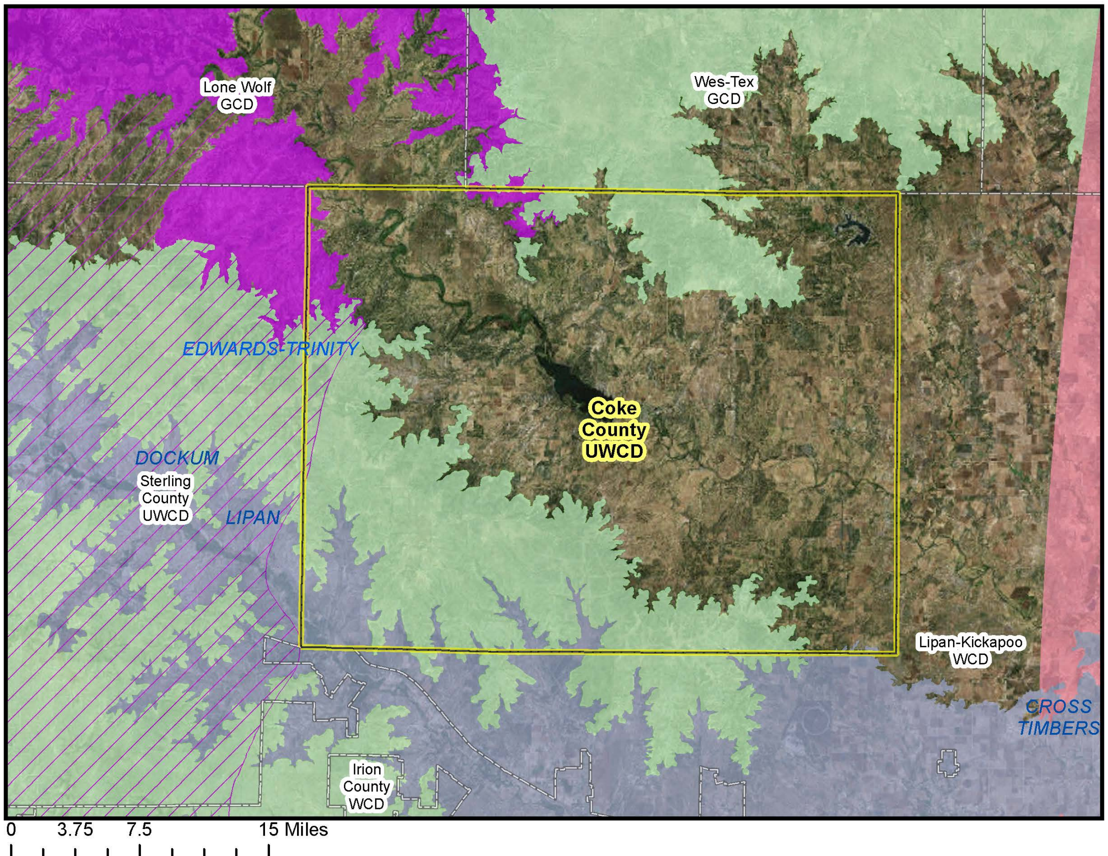


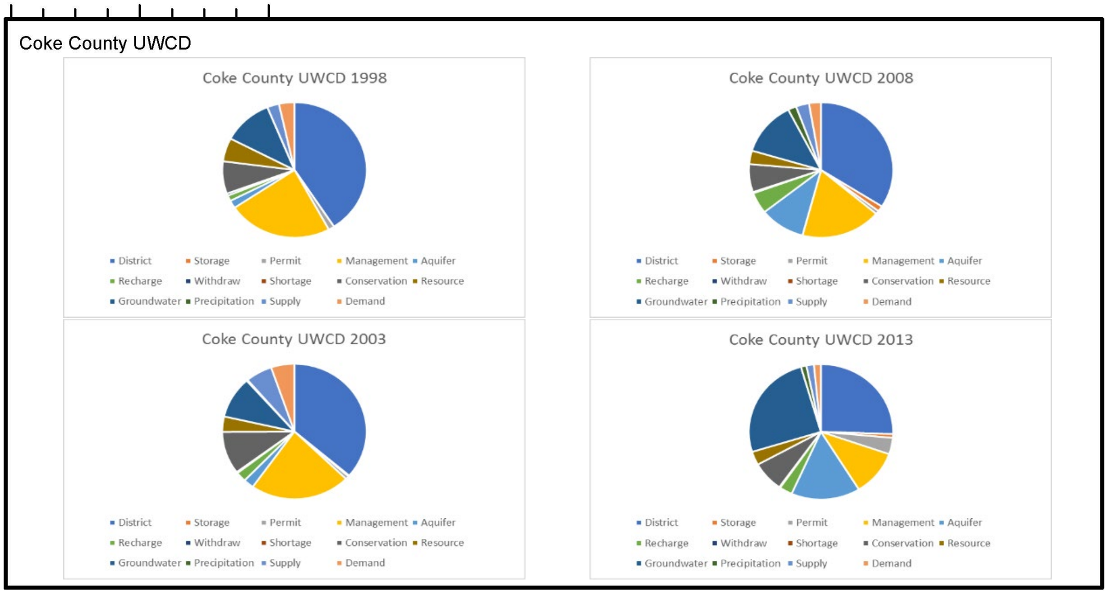
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



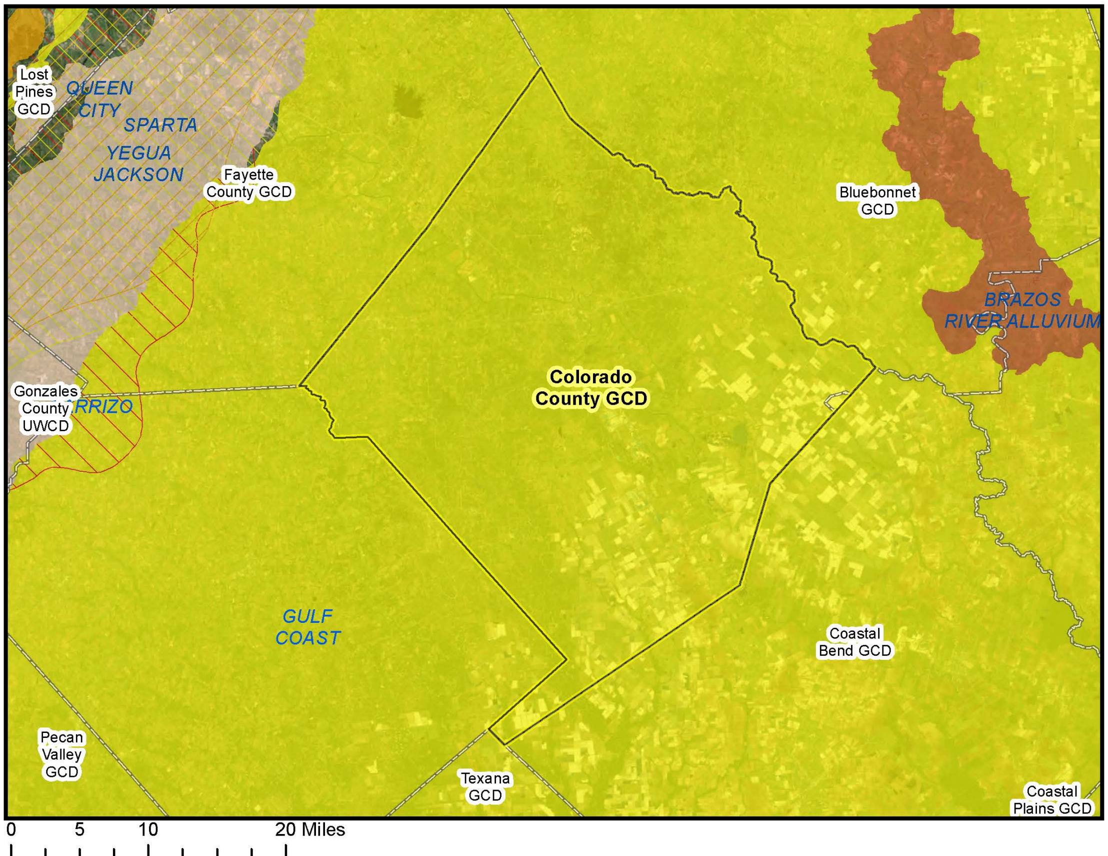


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

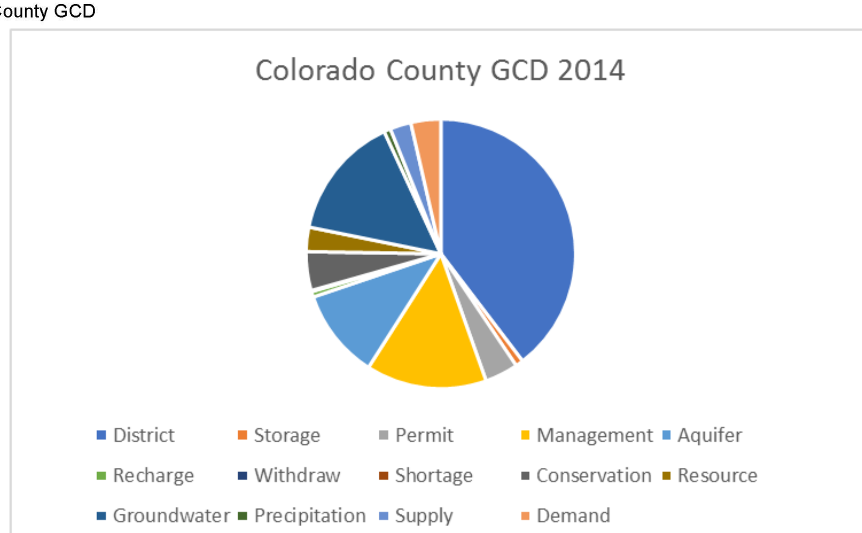




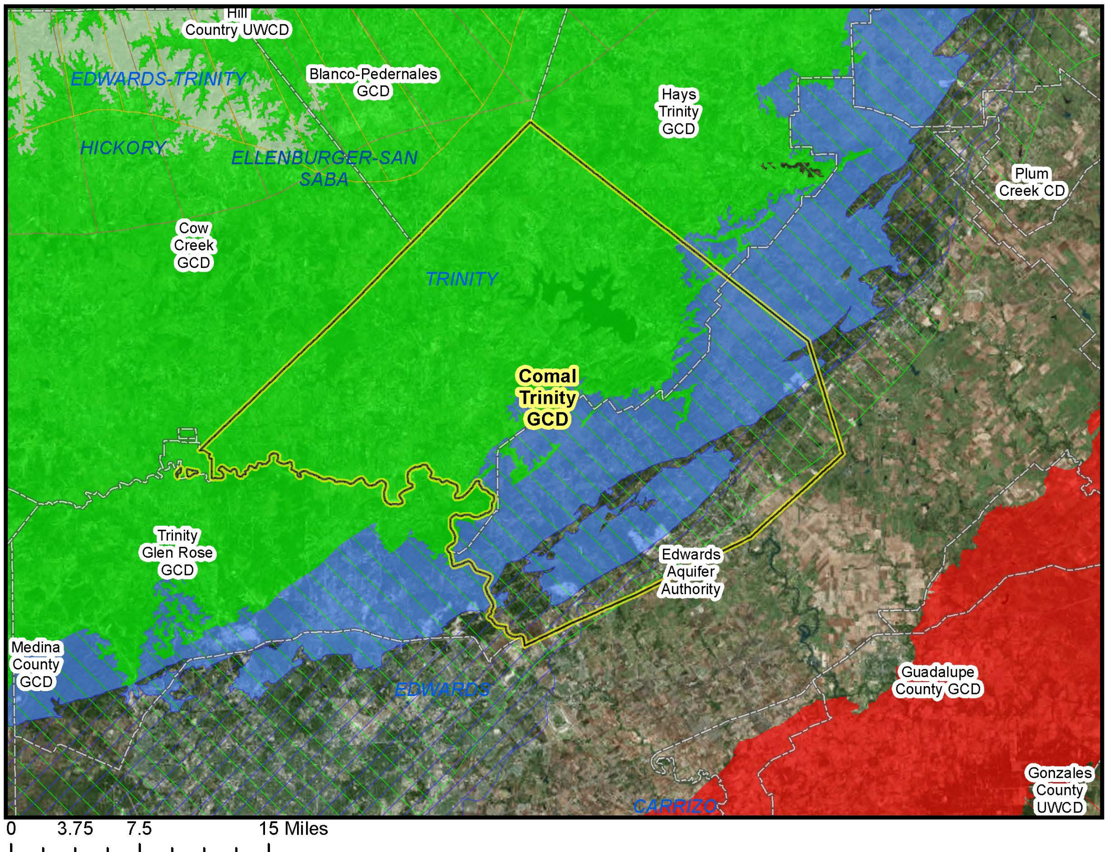
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



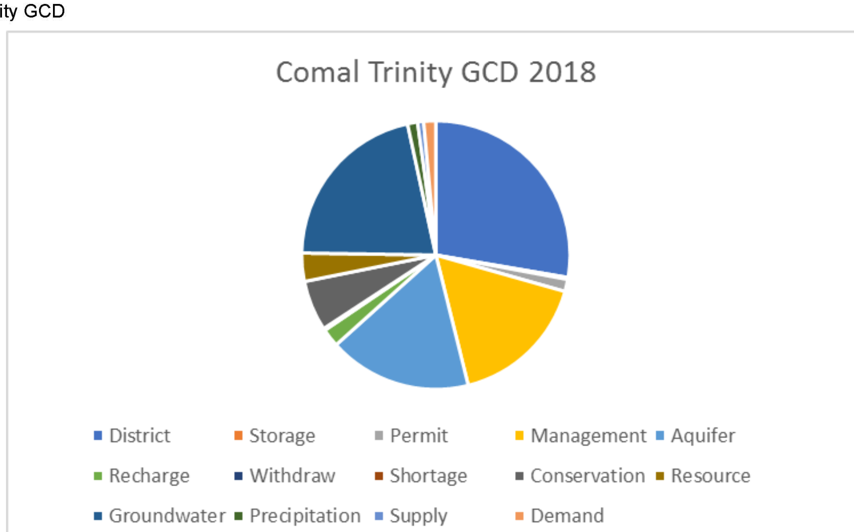
Colorado County GCD



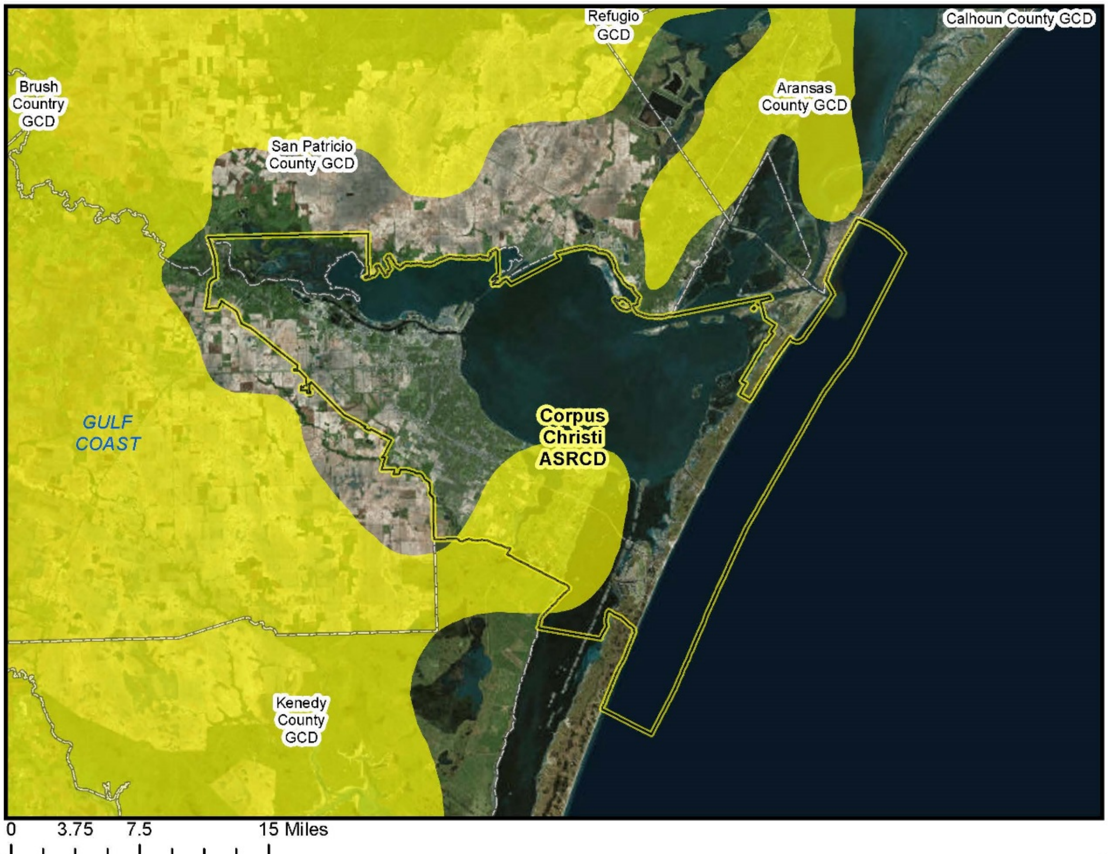
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



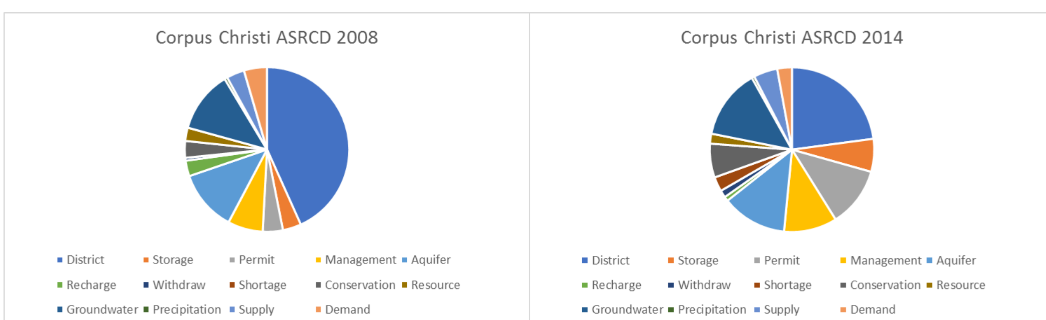
Comal Trinity GCD



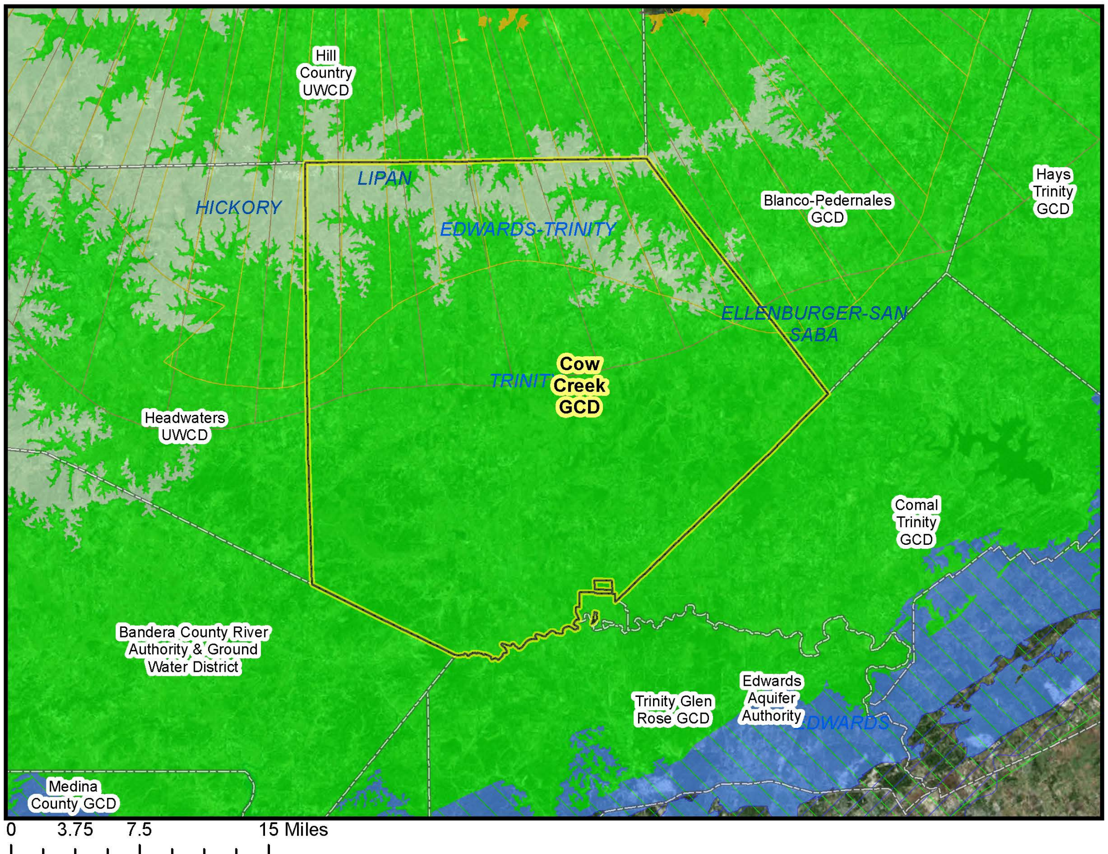
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

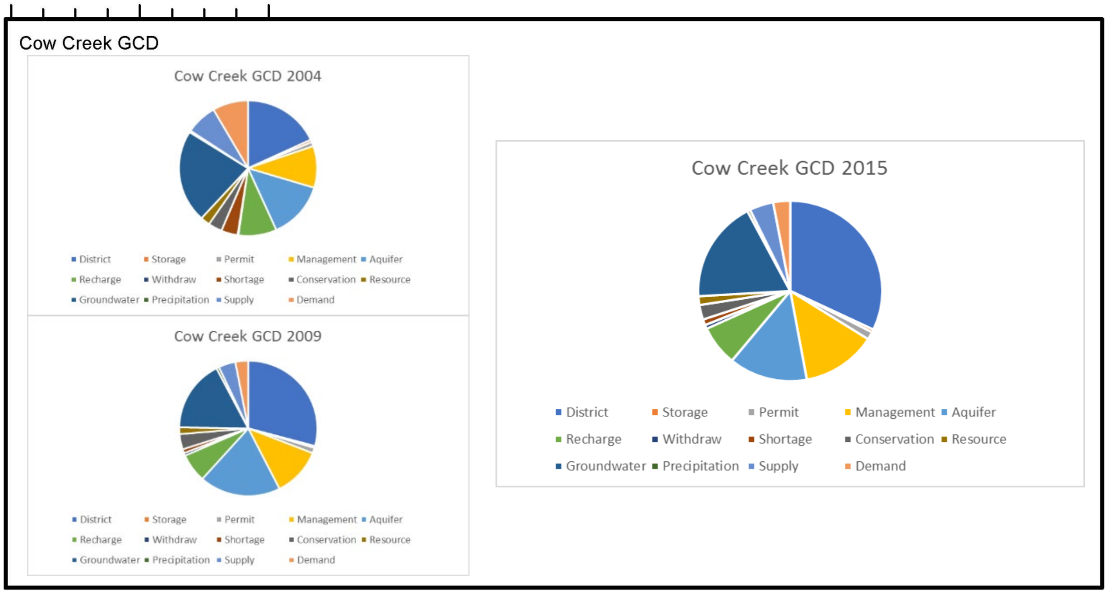


Corpus Christi ASRCD

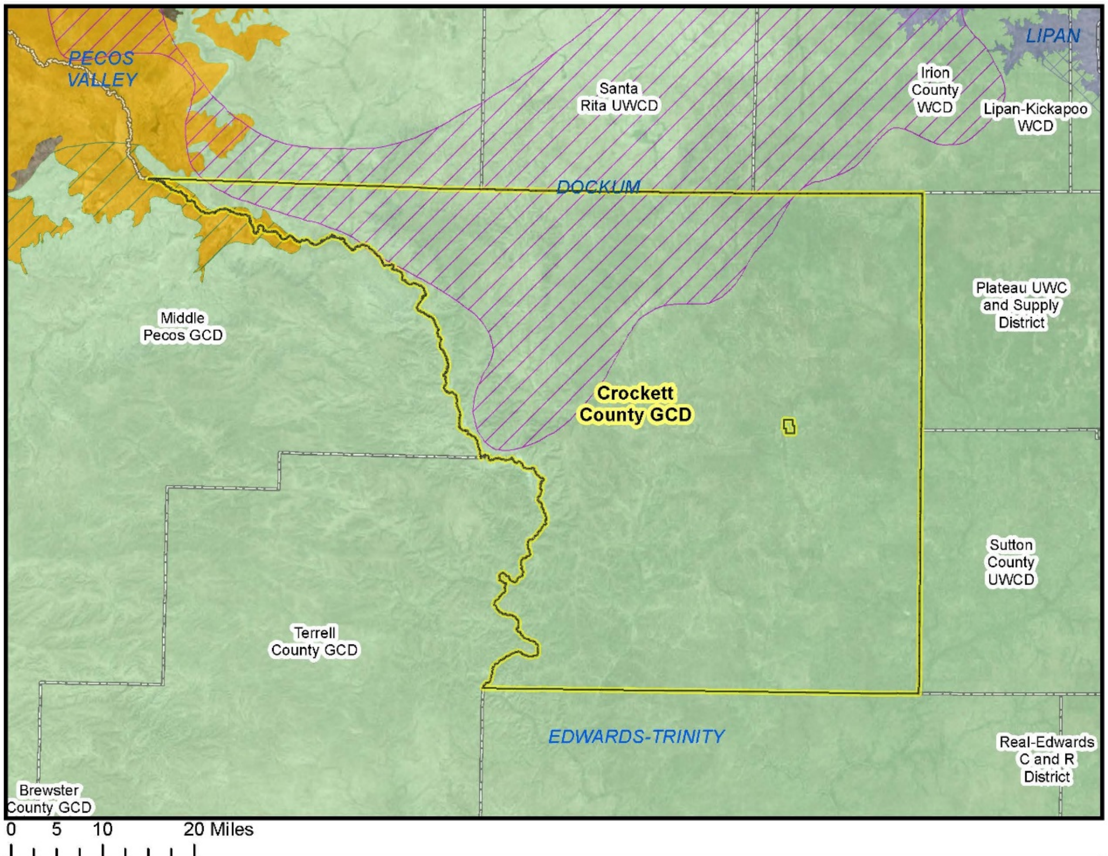


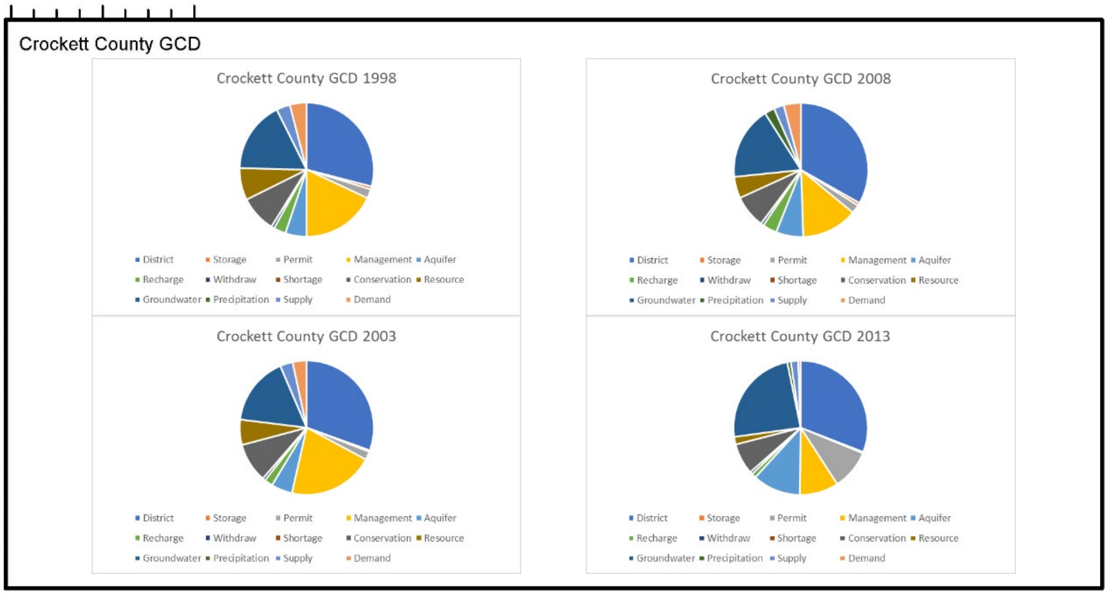
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



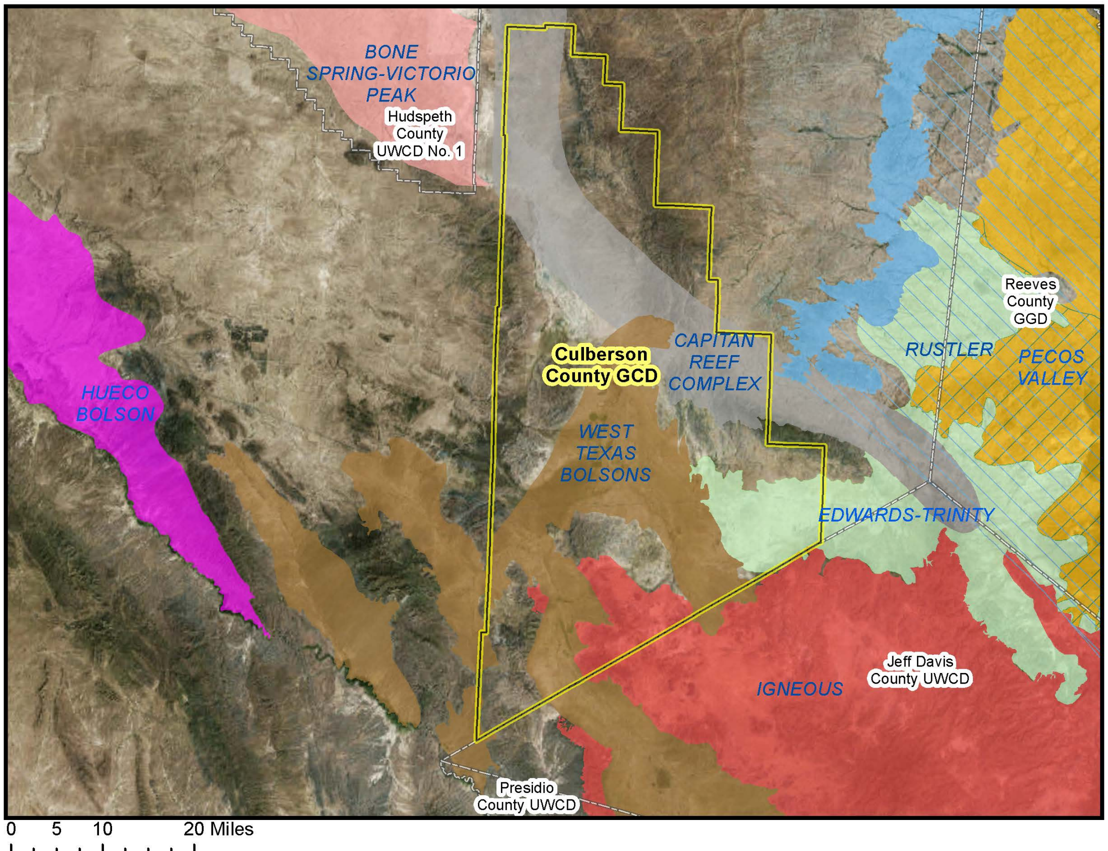


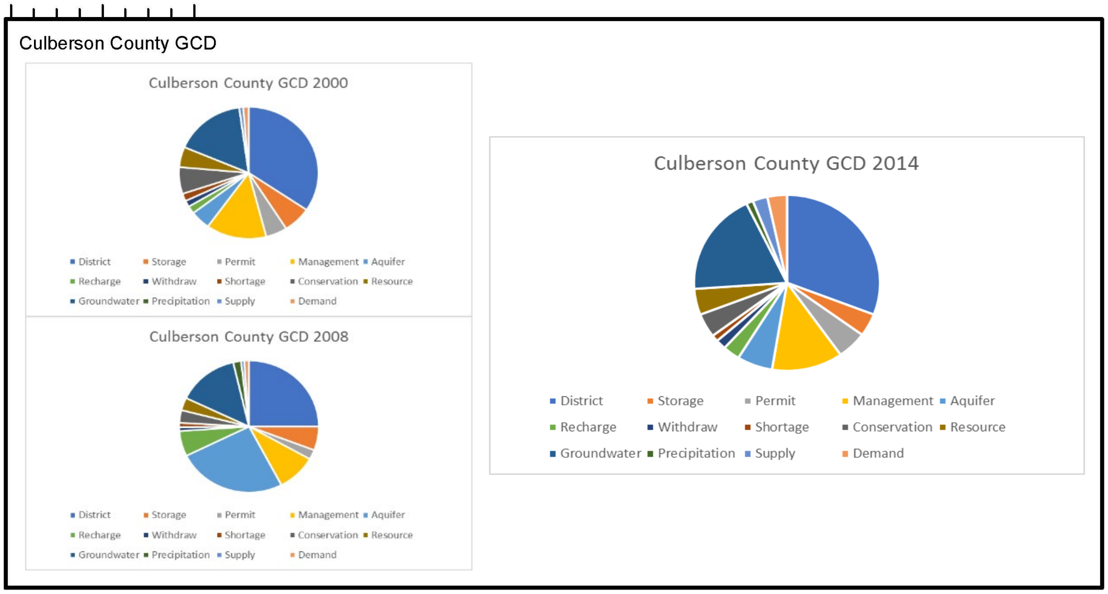
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



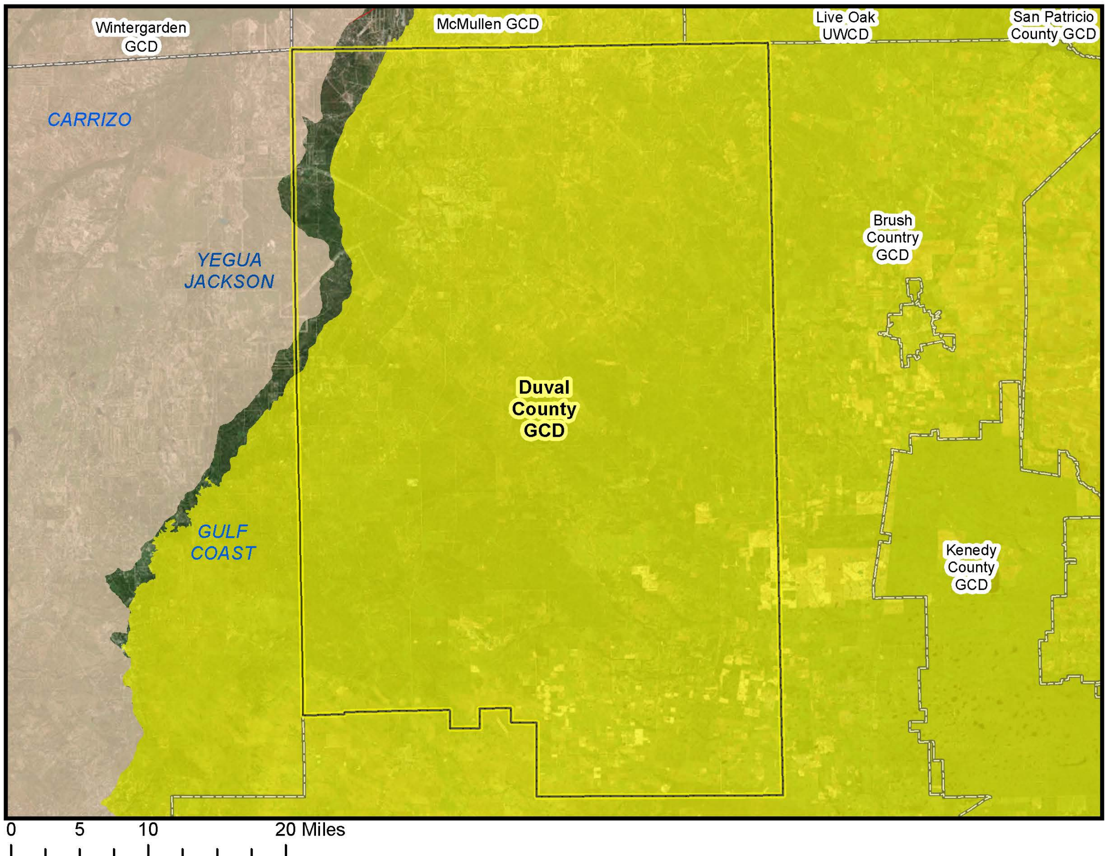


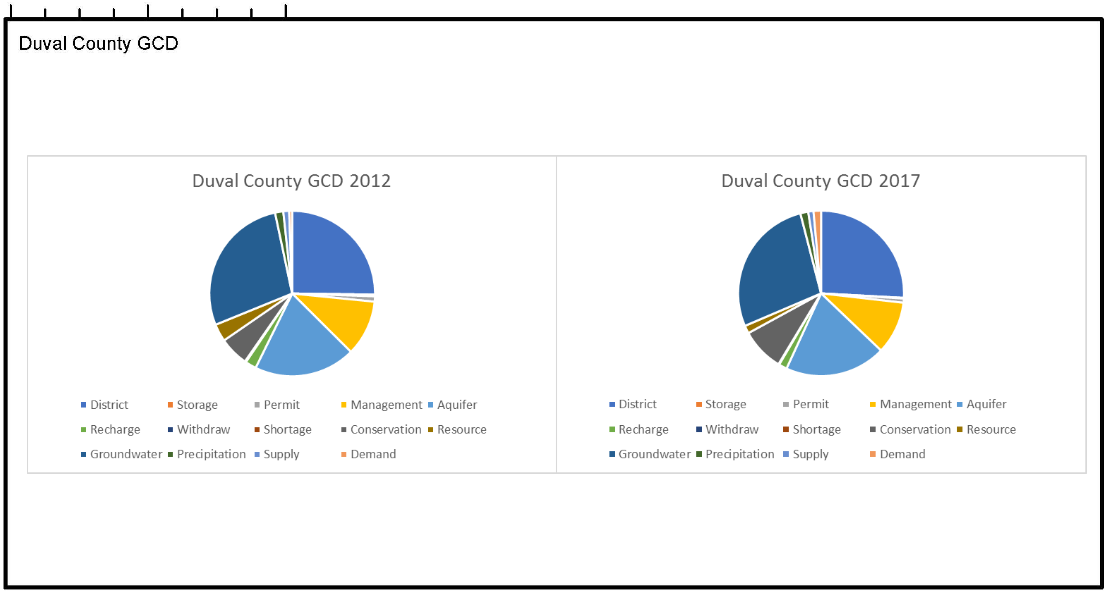
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



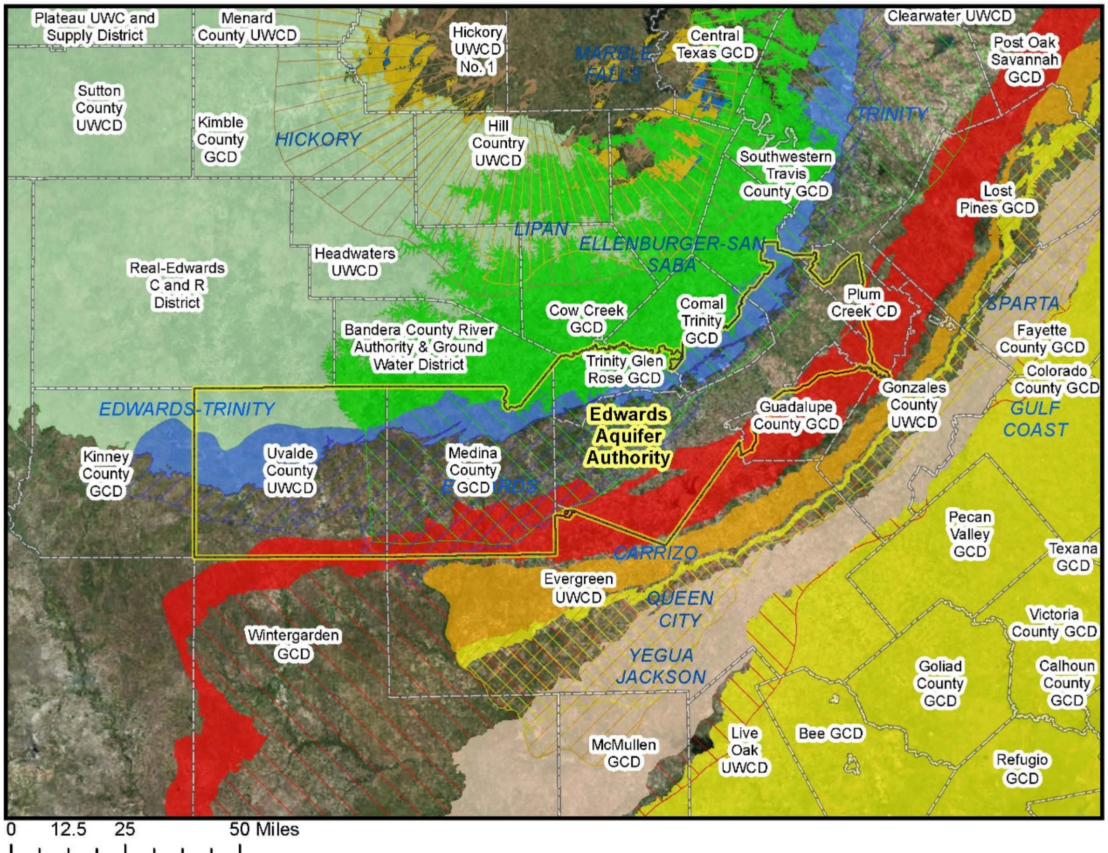


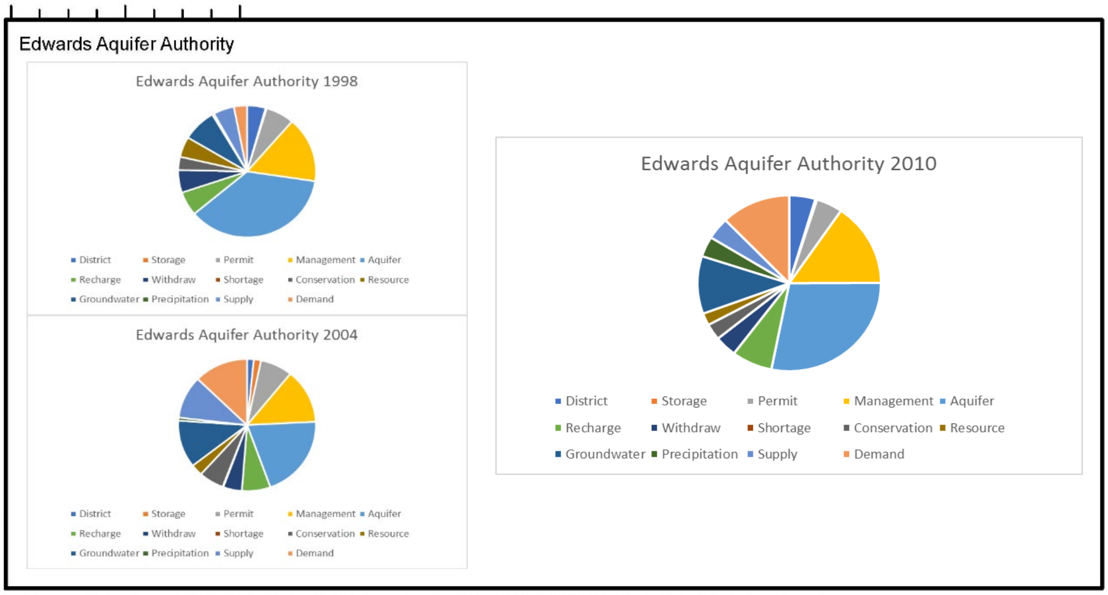
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



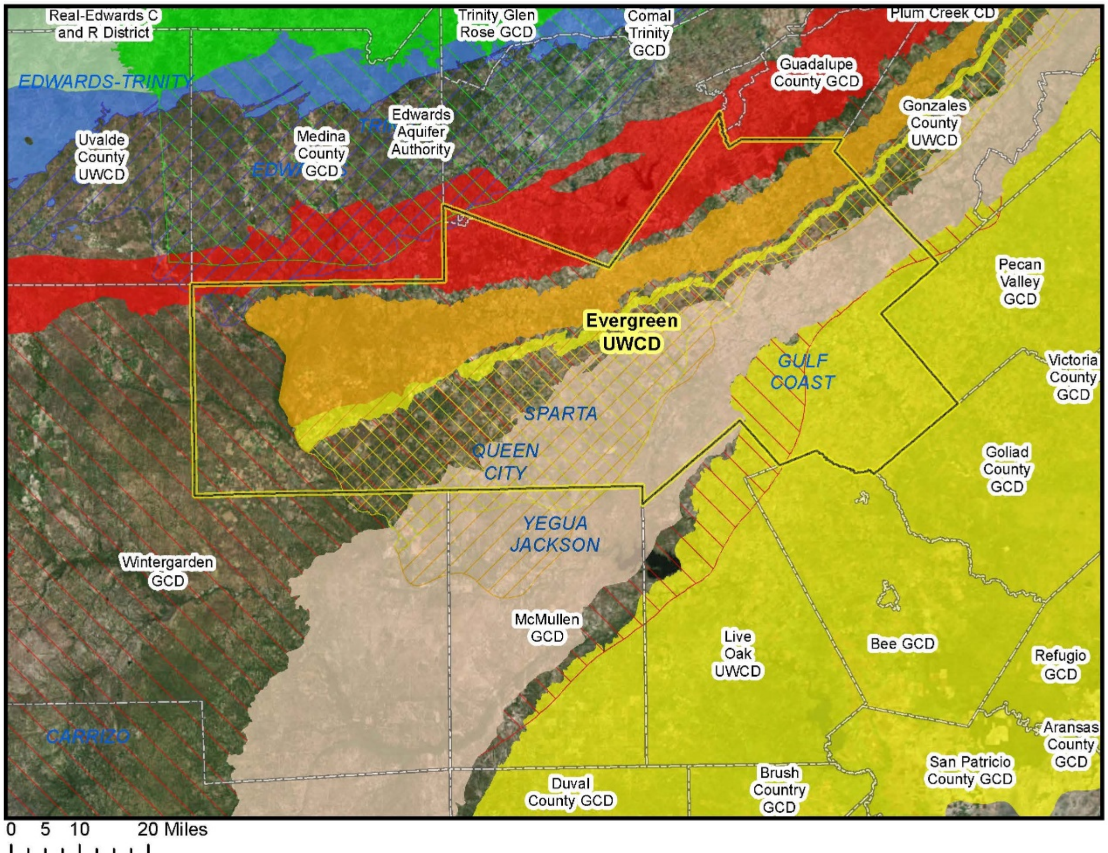


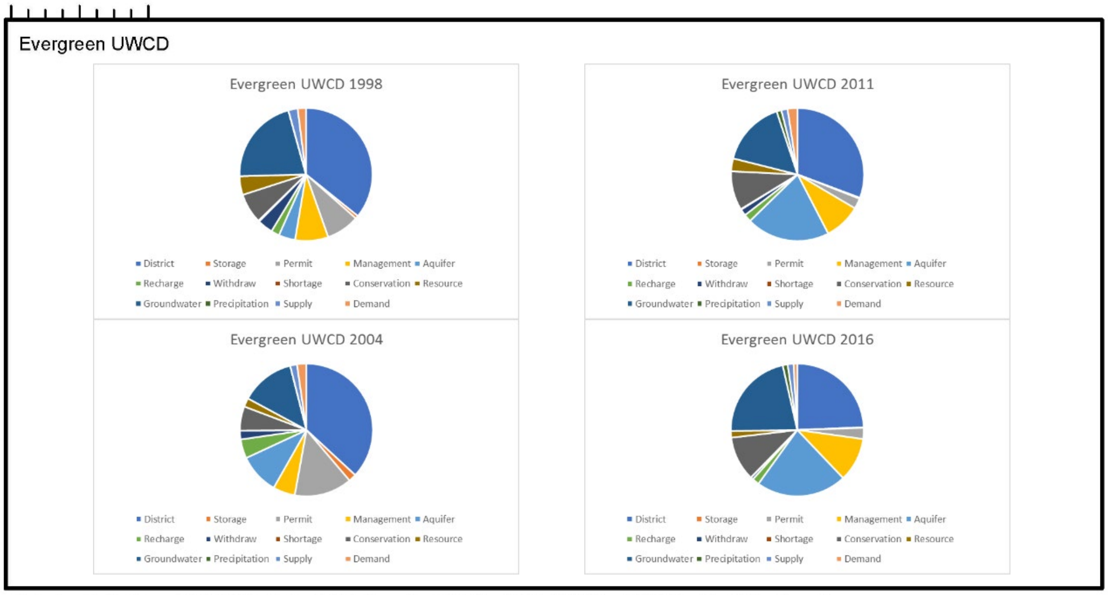
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



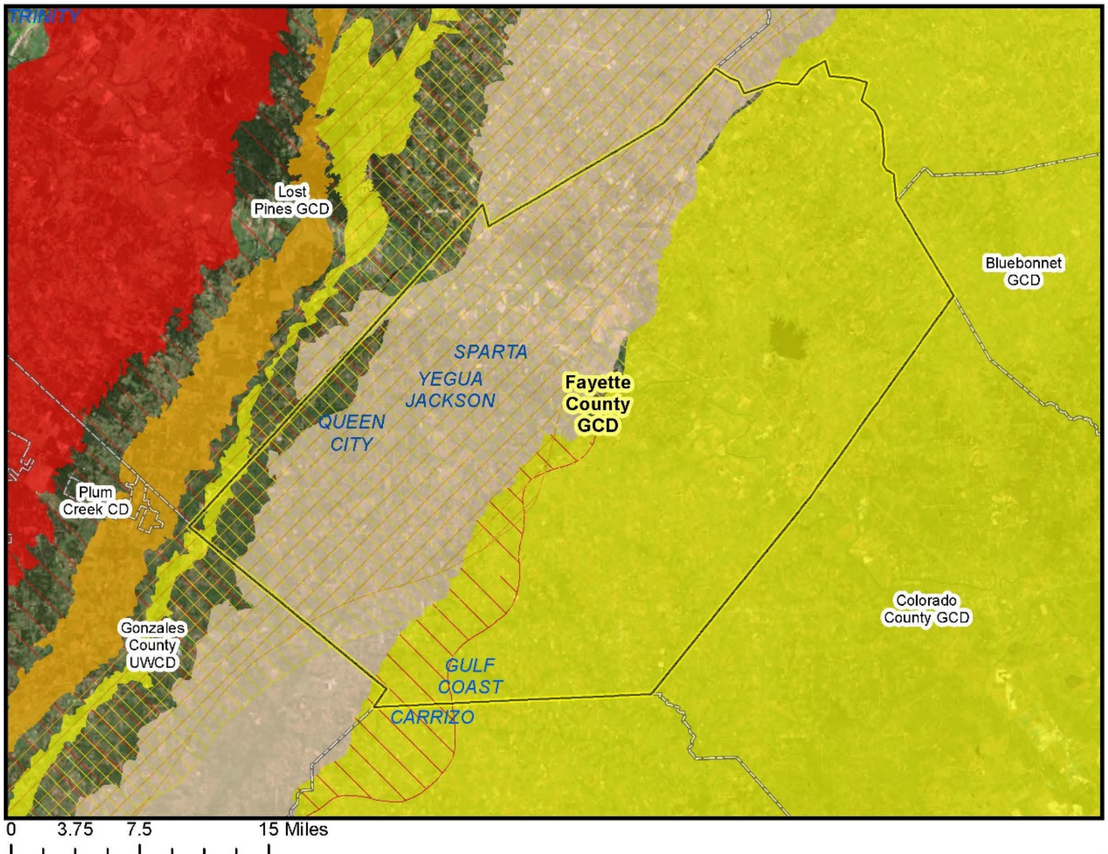


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

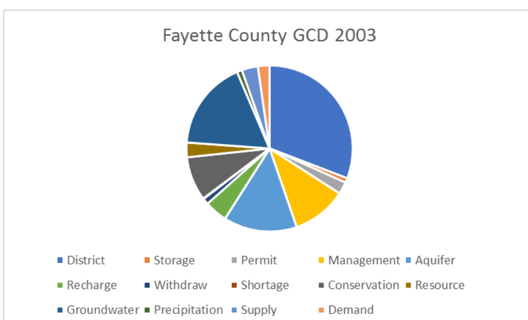


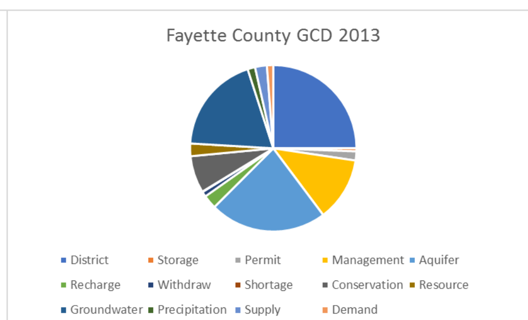


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

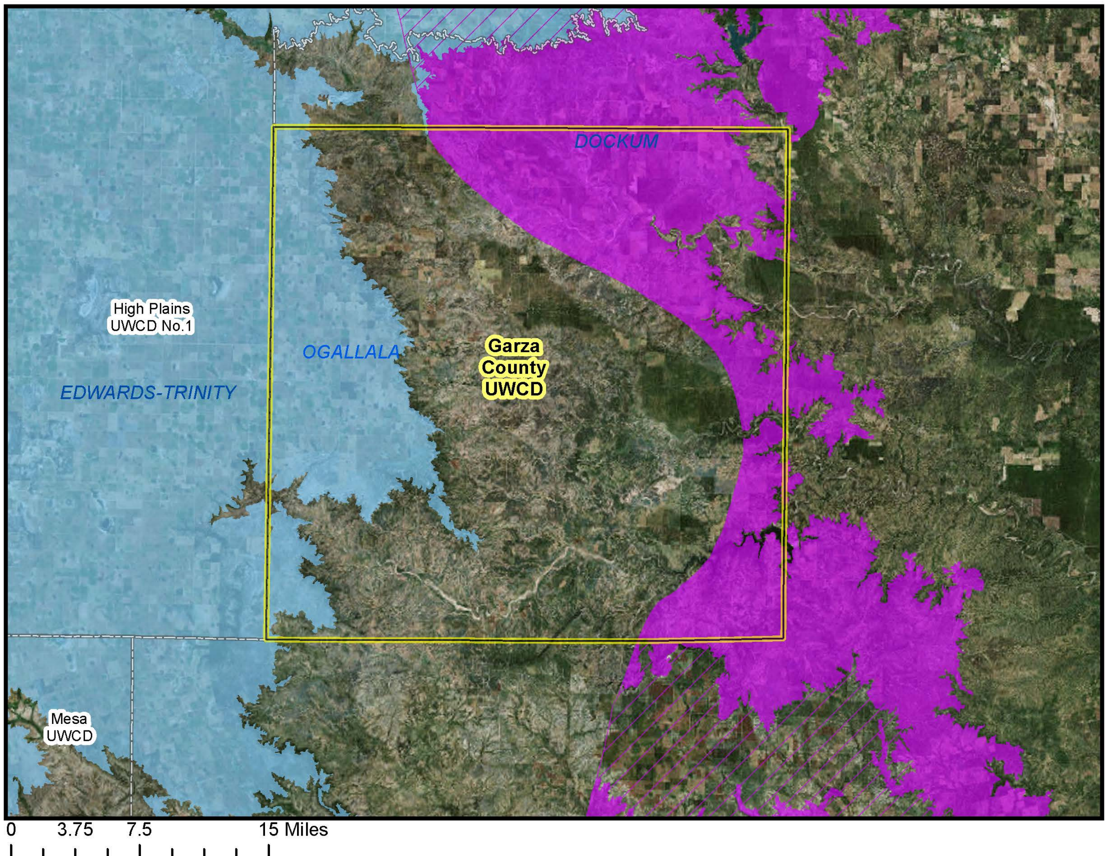


Fayette County GCD

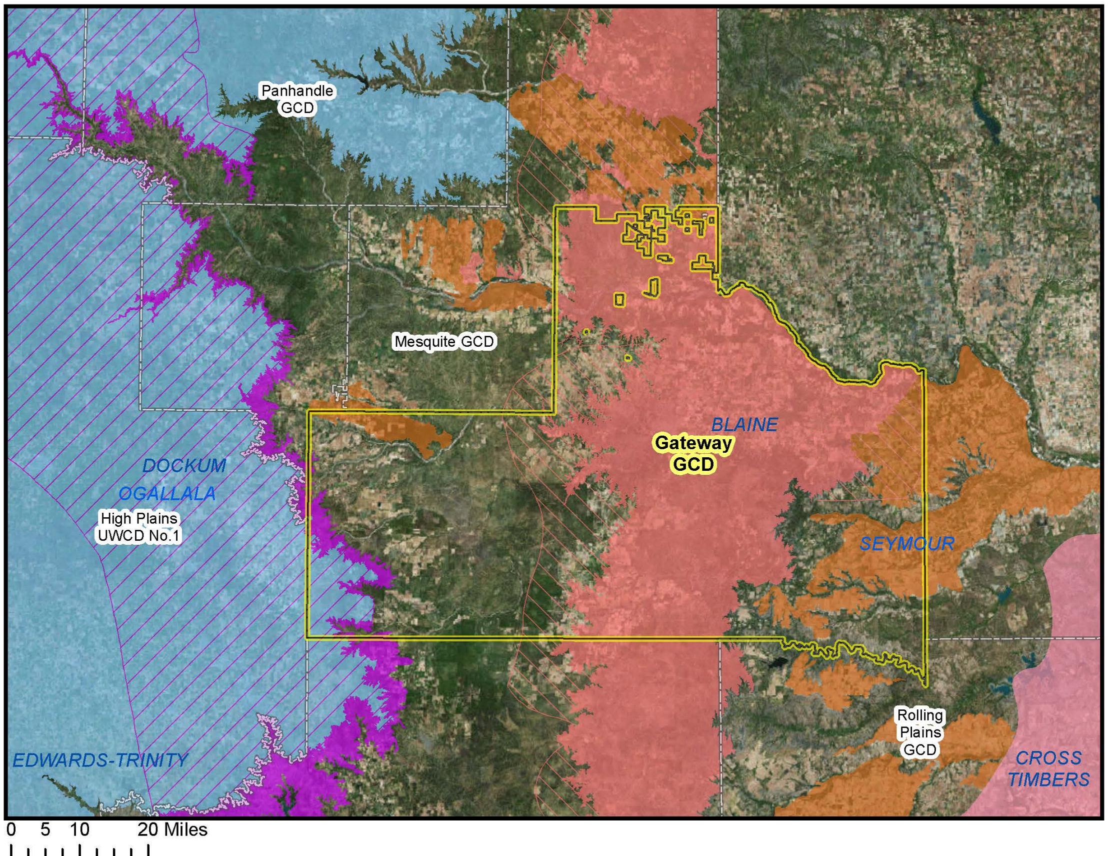




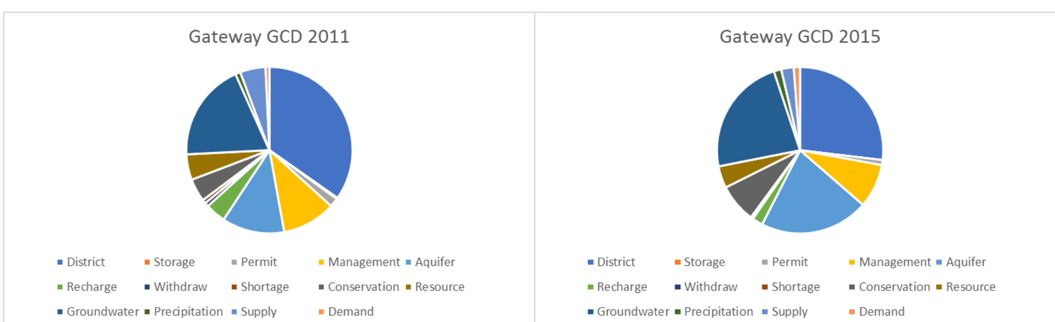
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



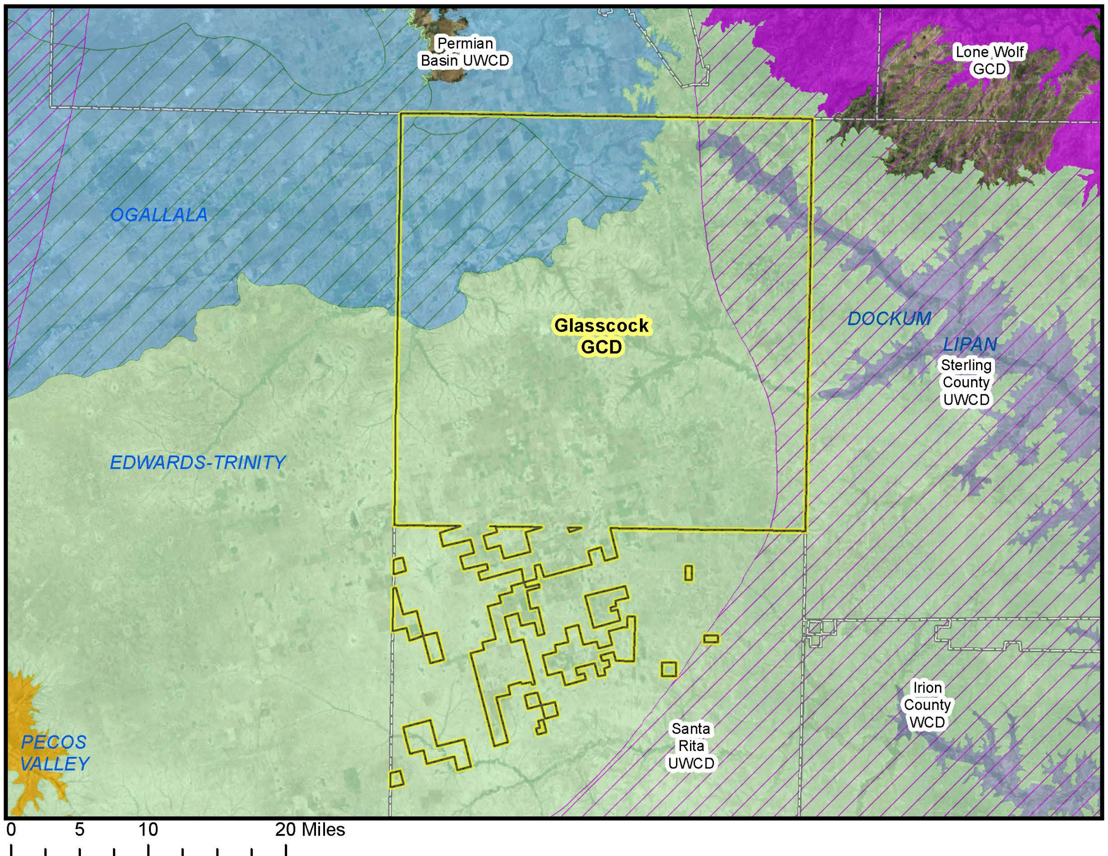
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

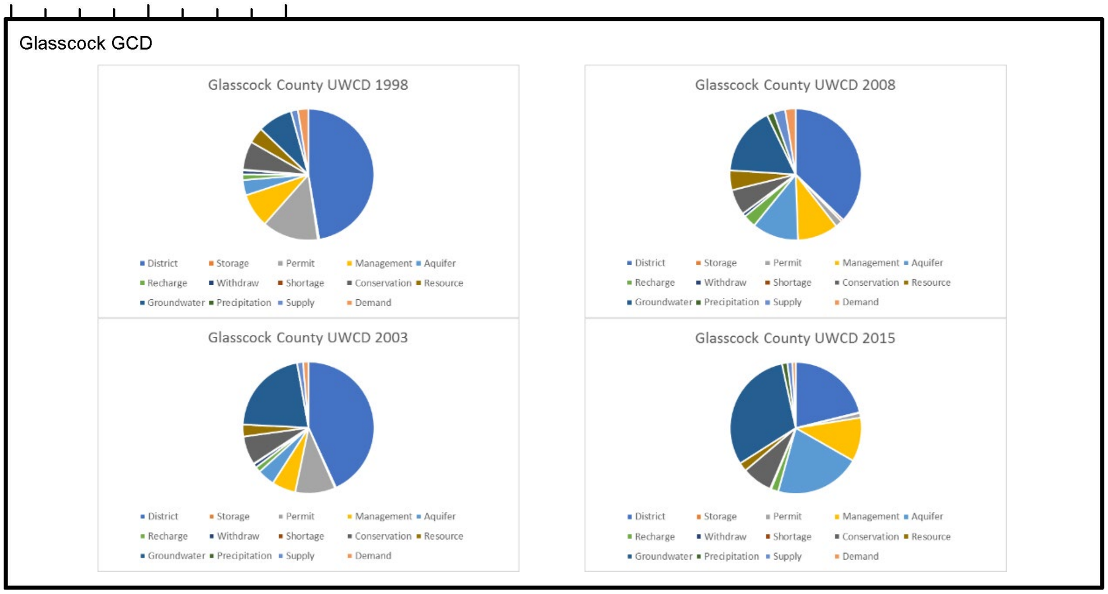


Gateway GCD

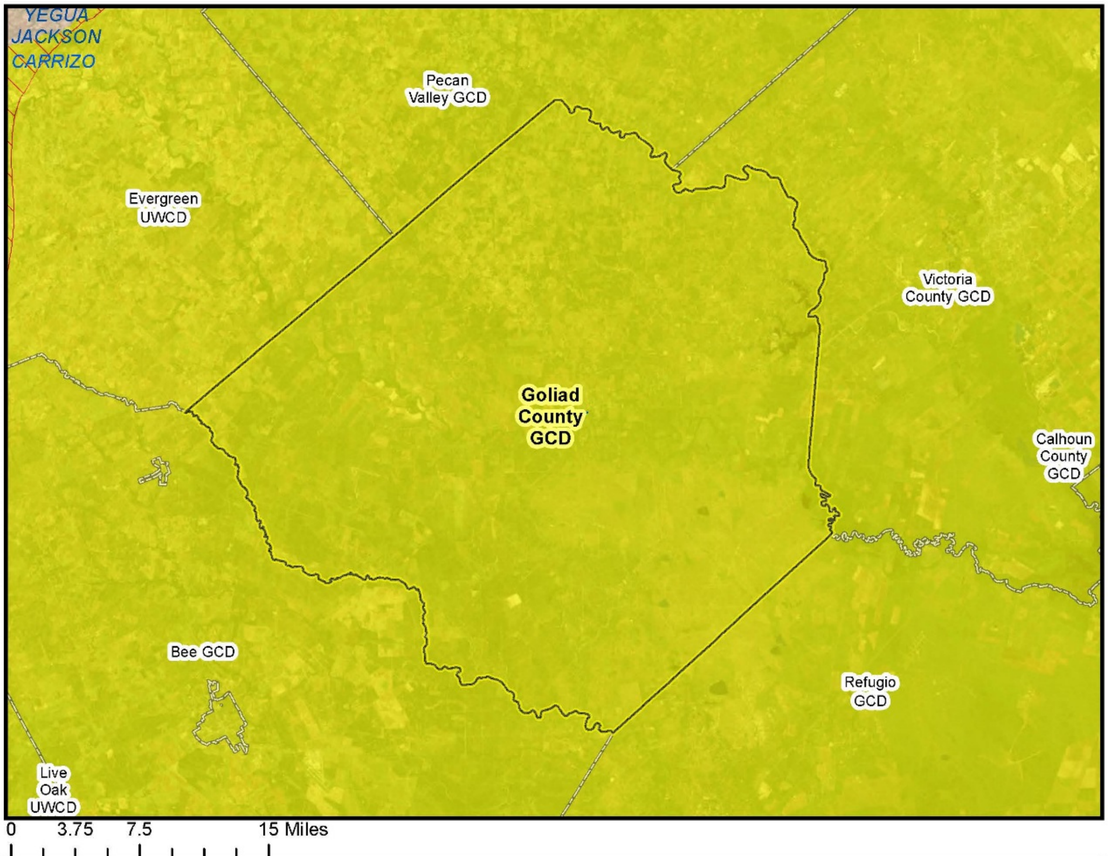


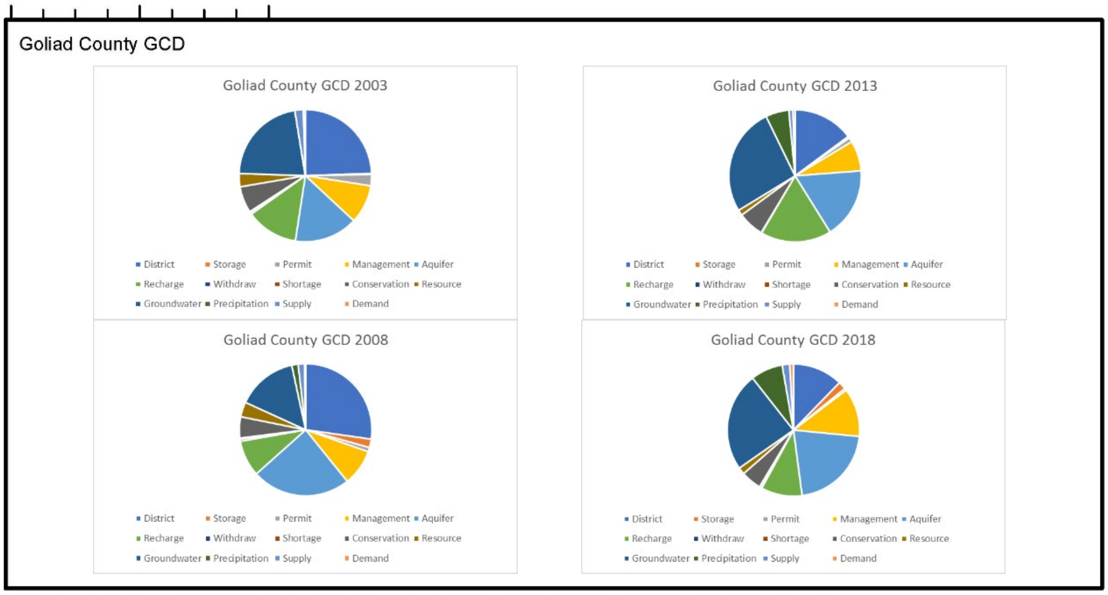
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



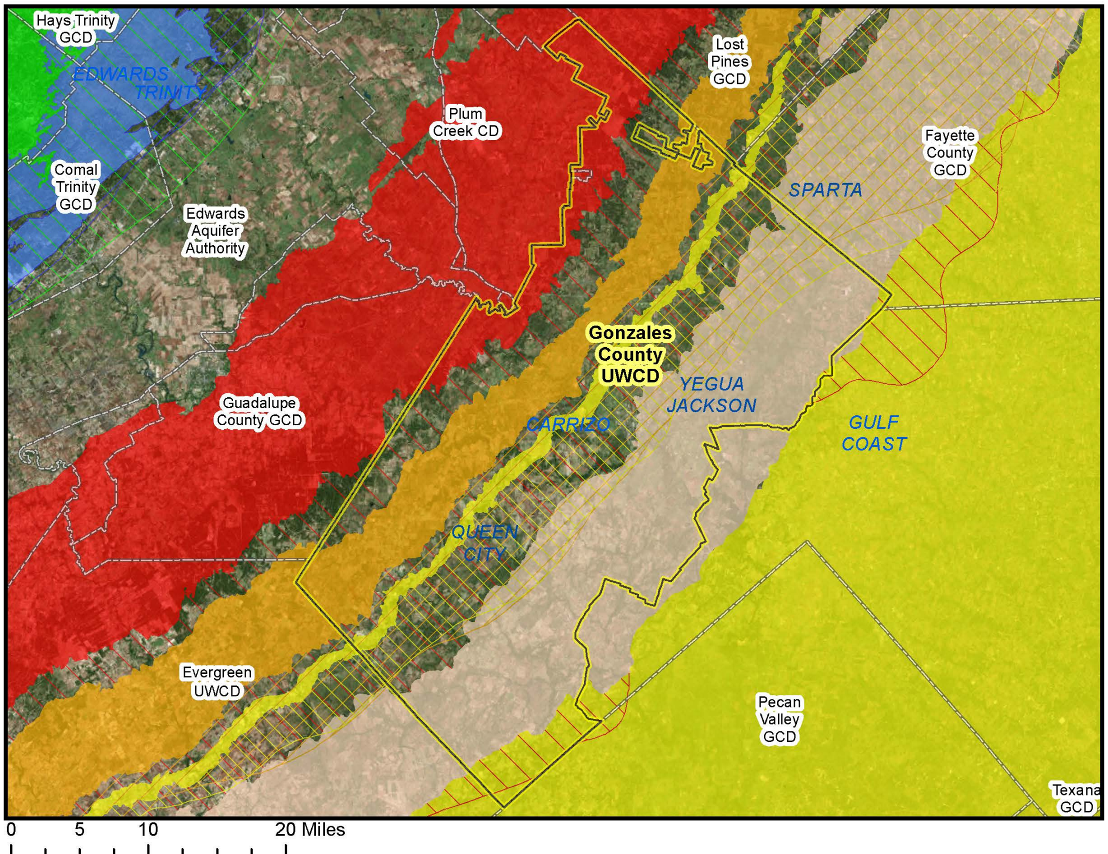


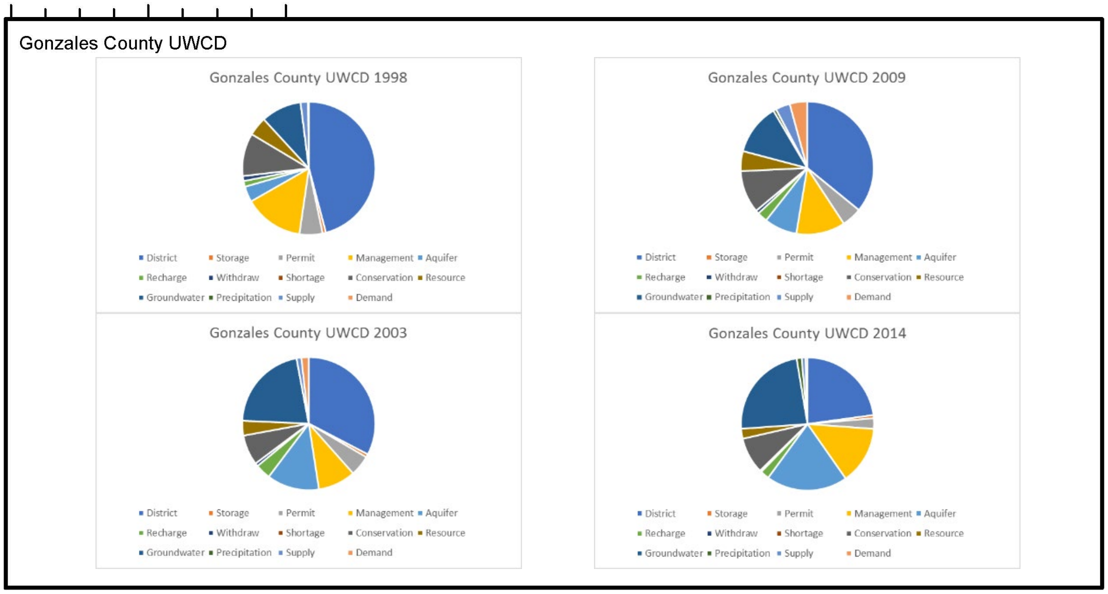
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



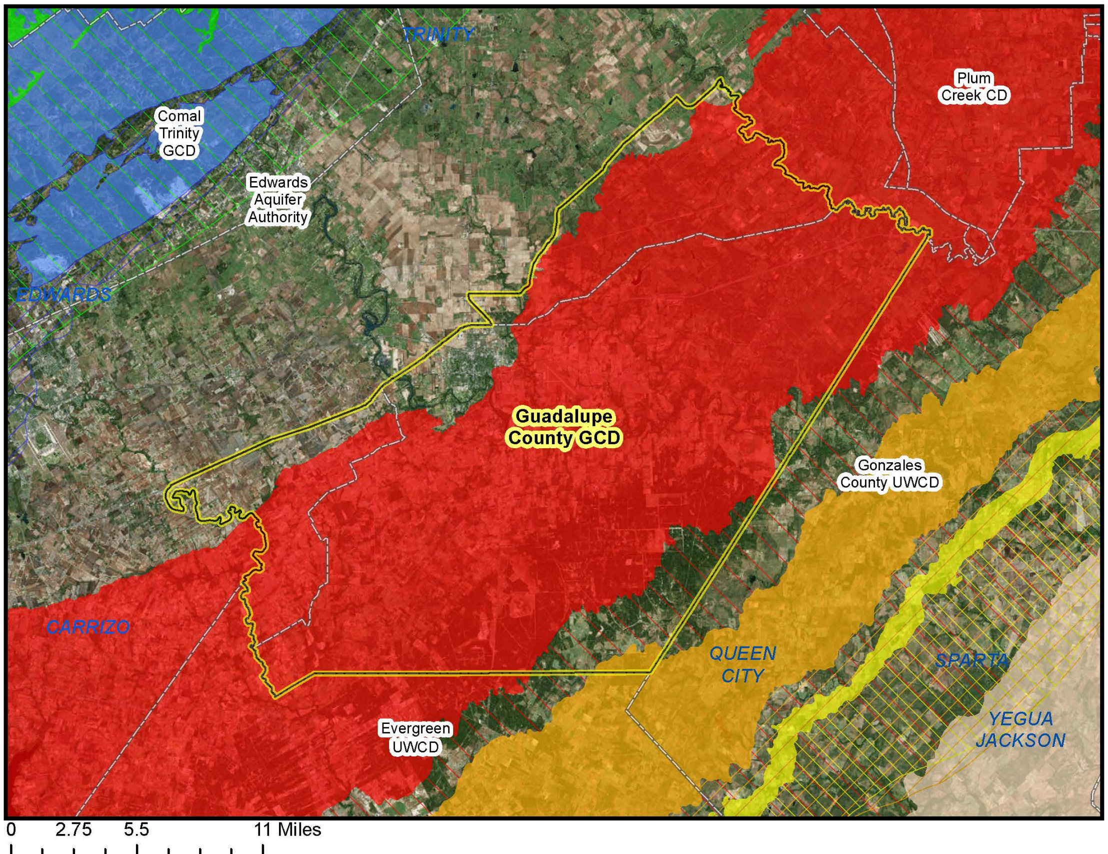


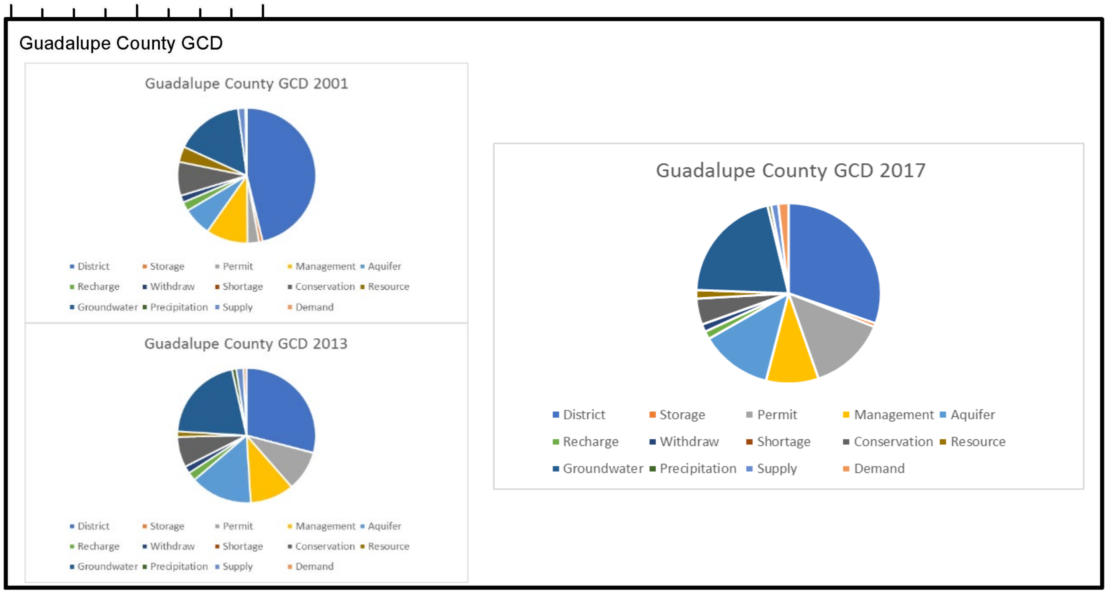
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



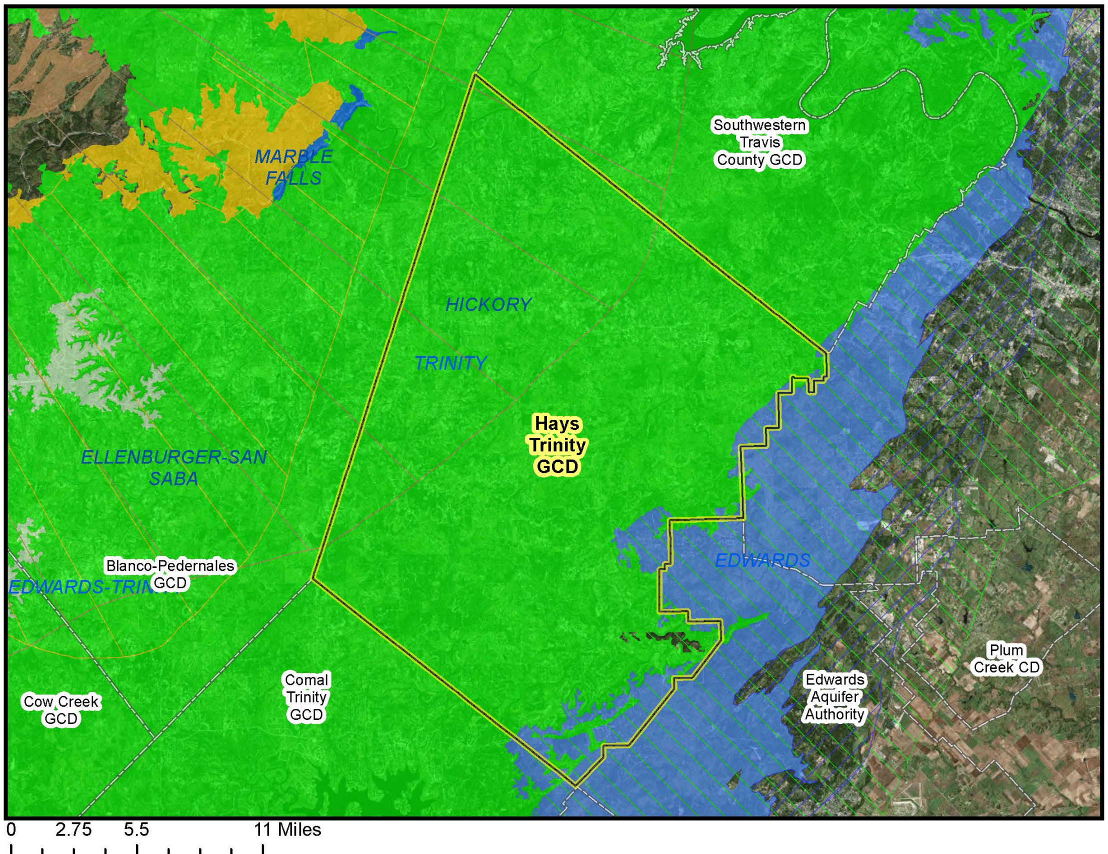


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

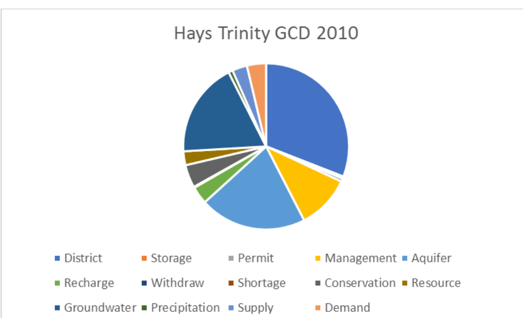


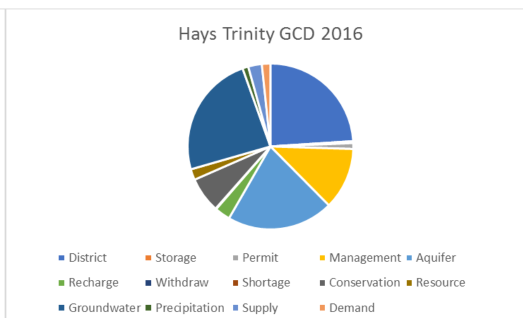


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

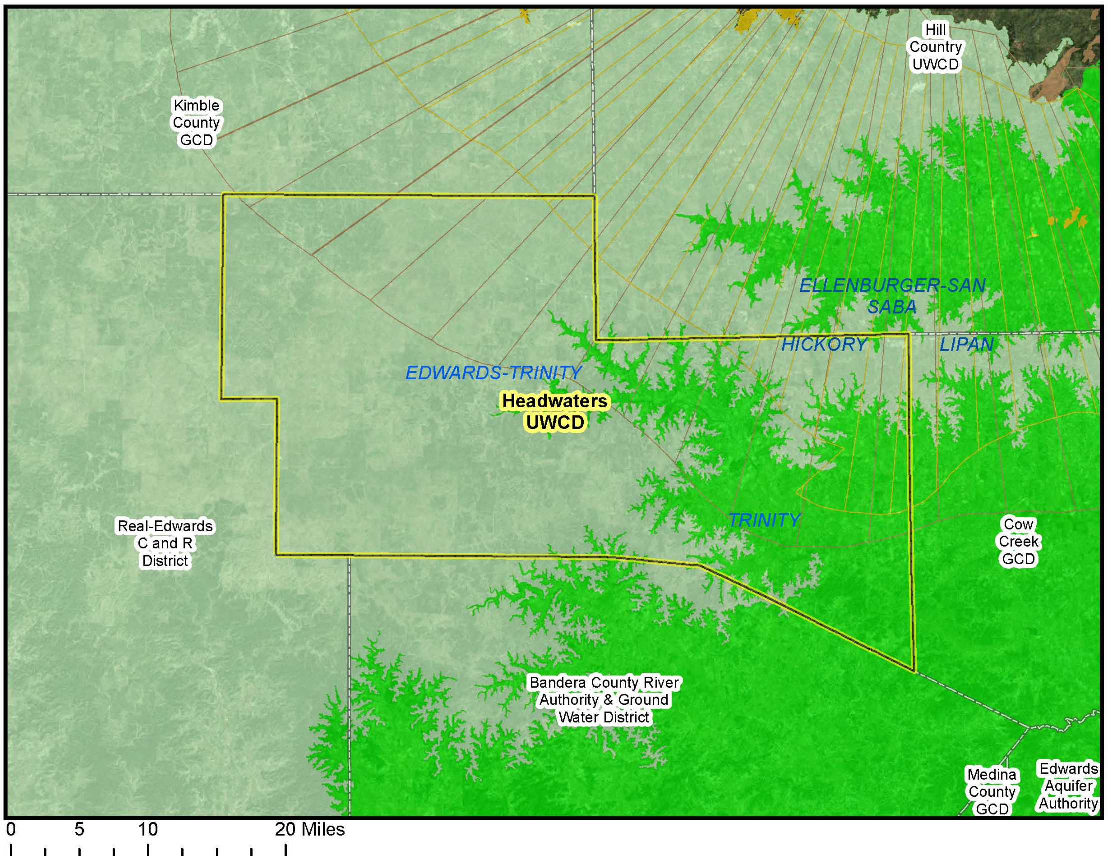


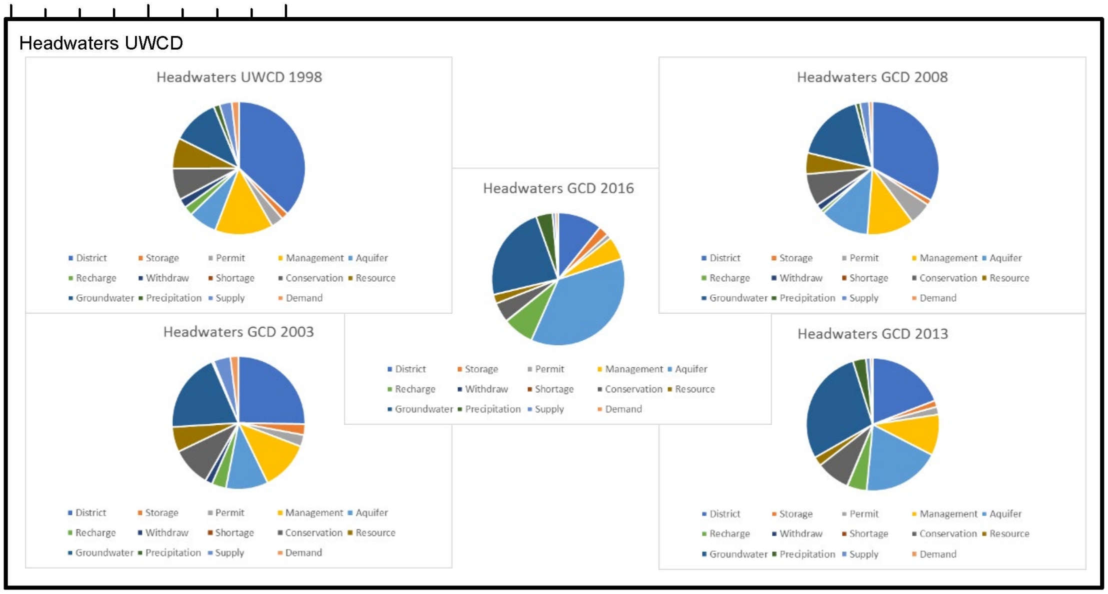
Hays Trinity GCD



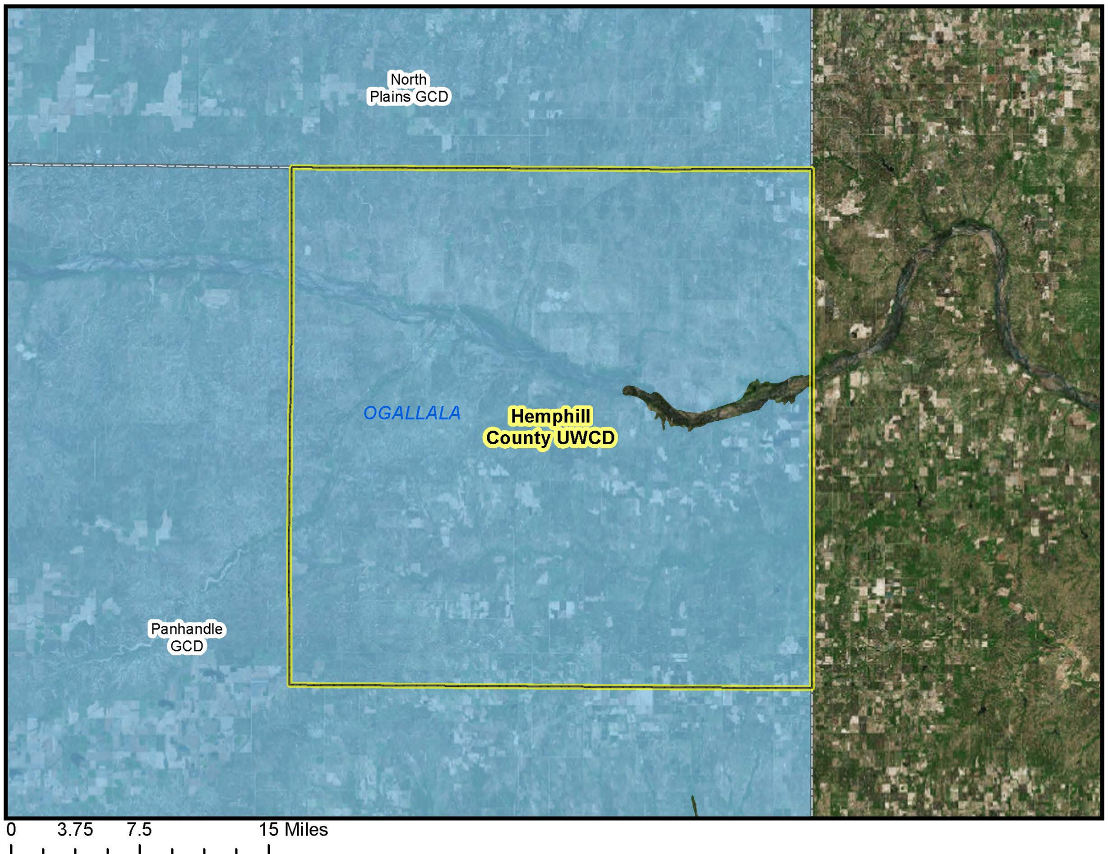


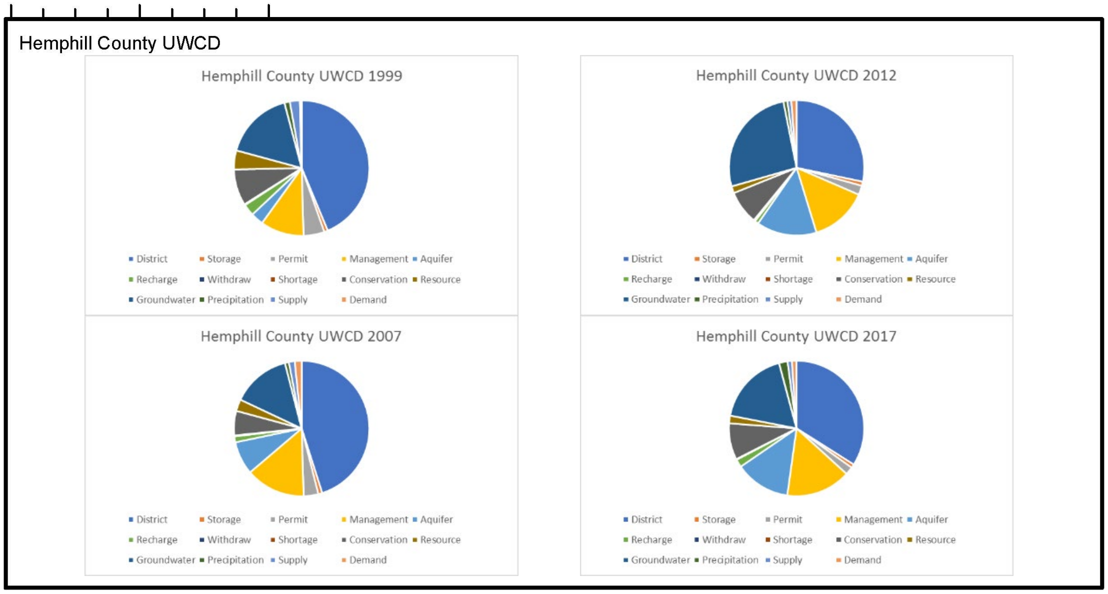
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



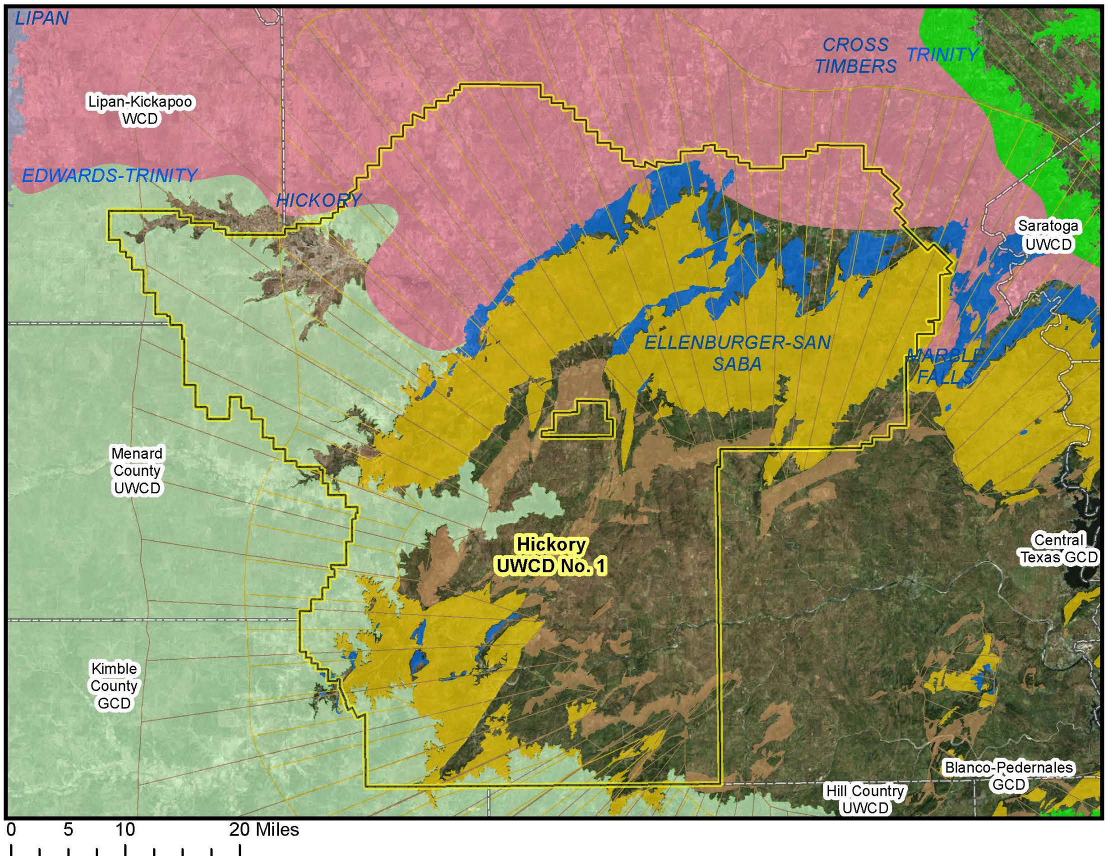


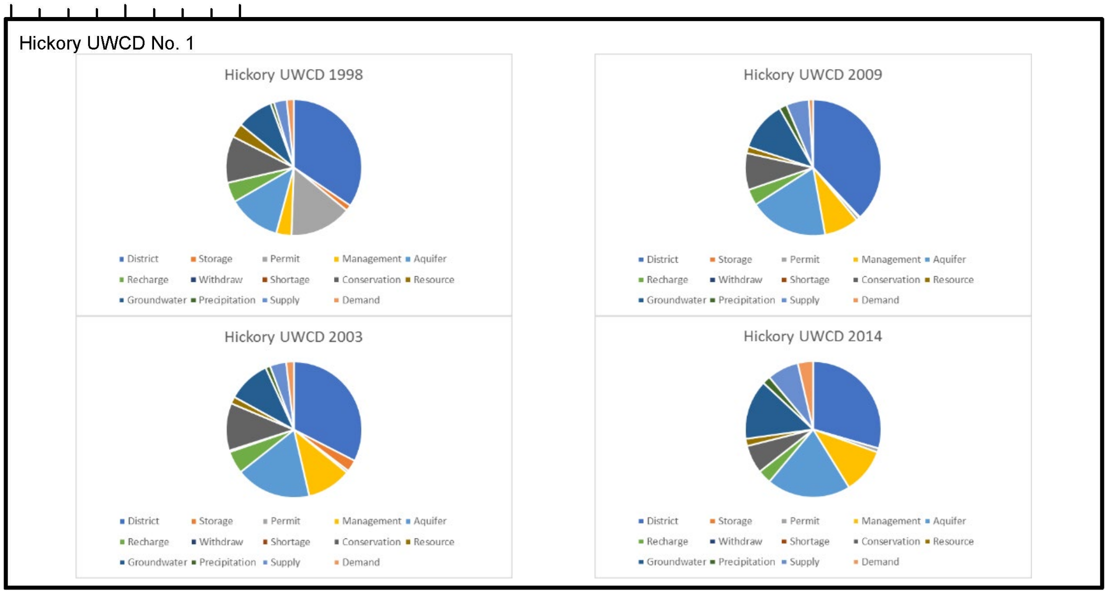
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



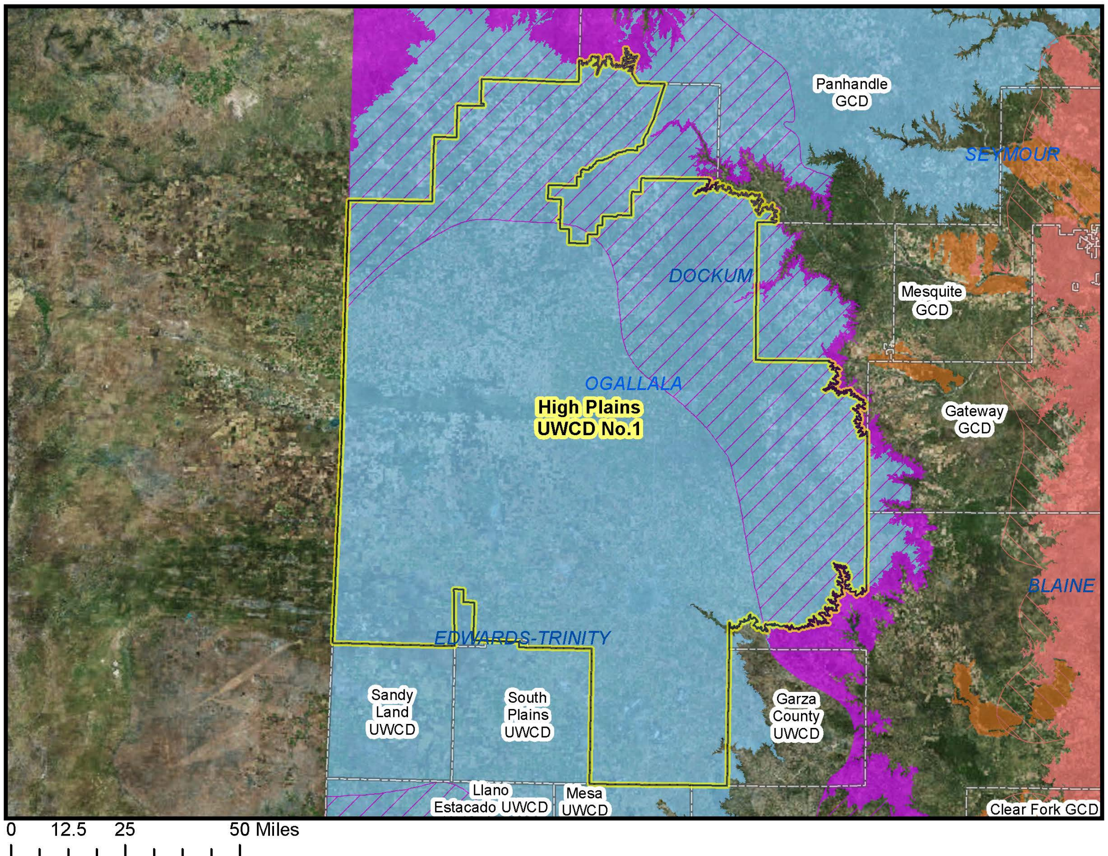


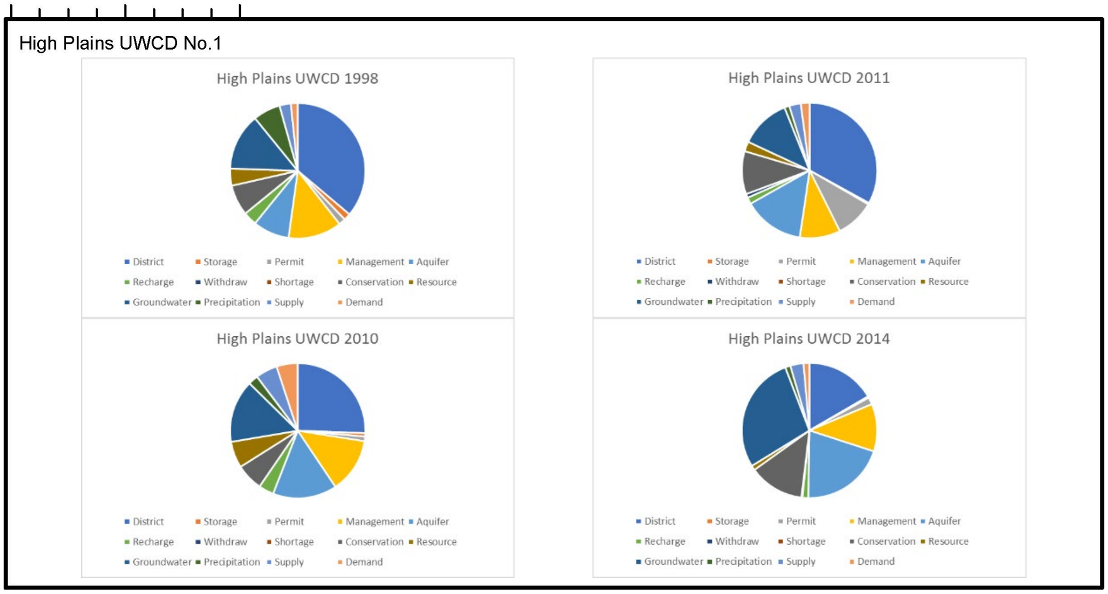
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



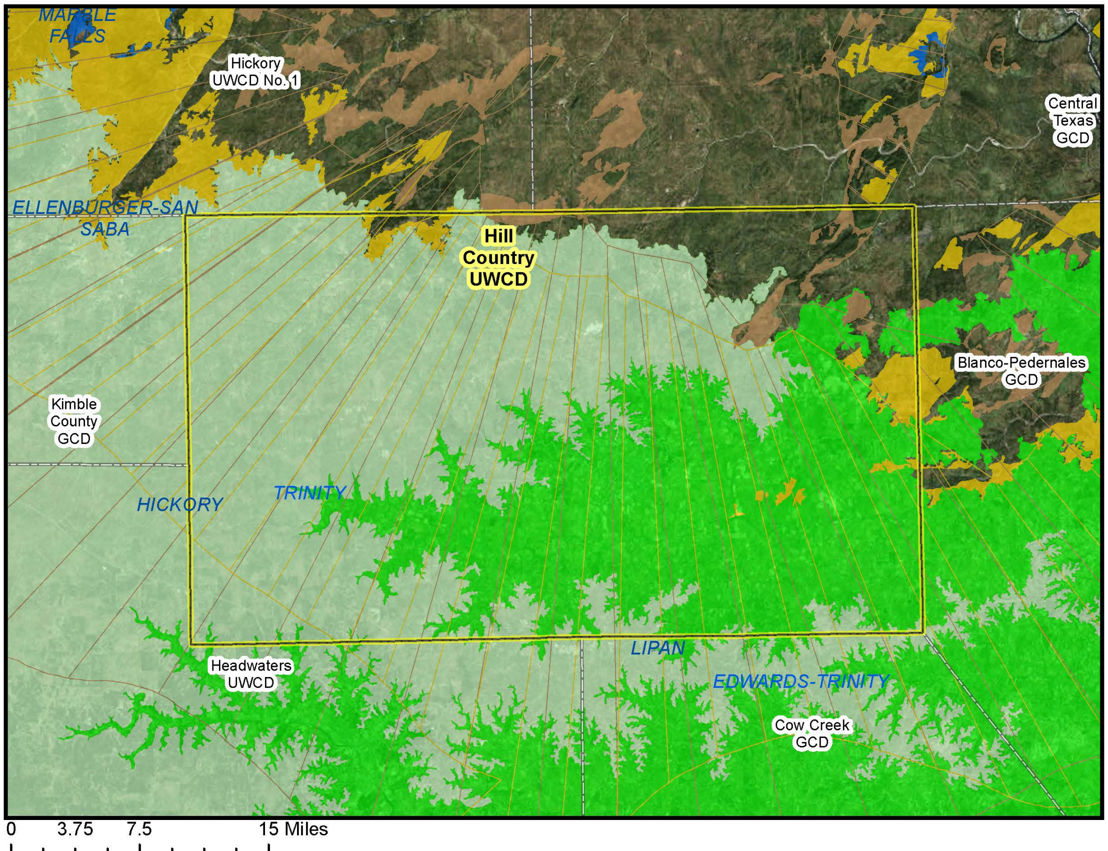


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

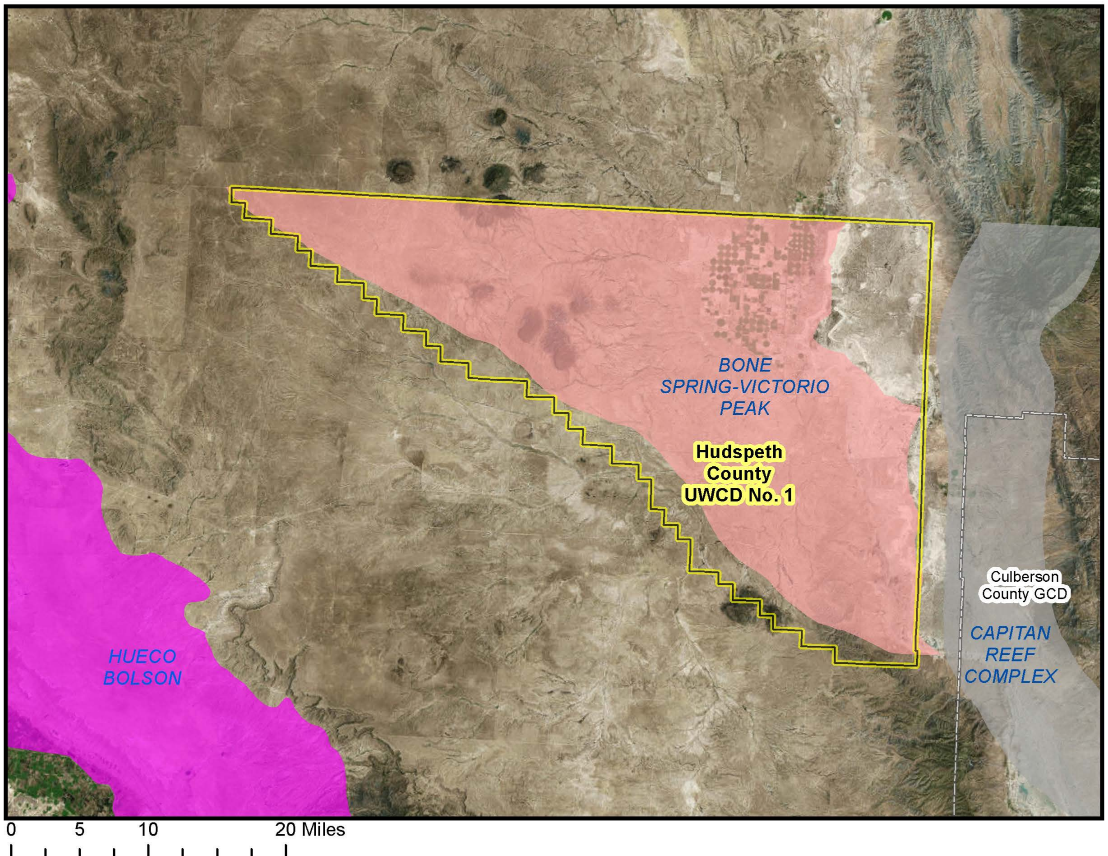


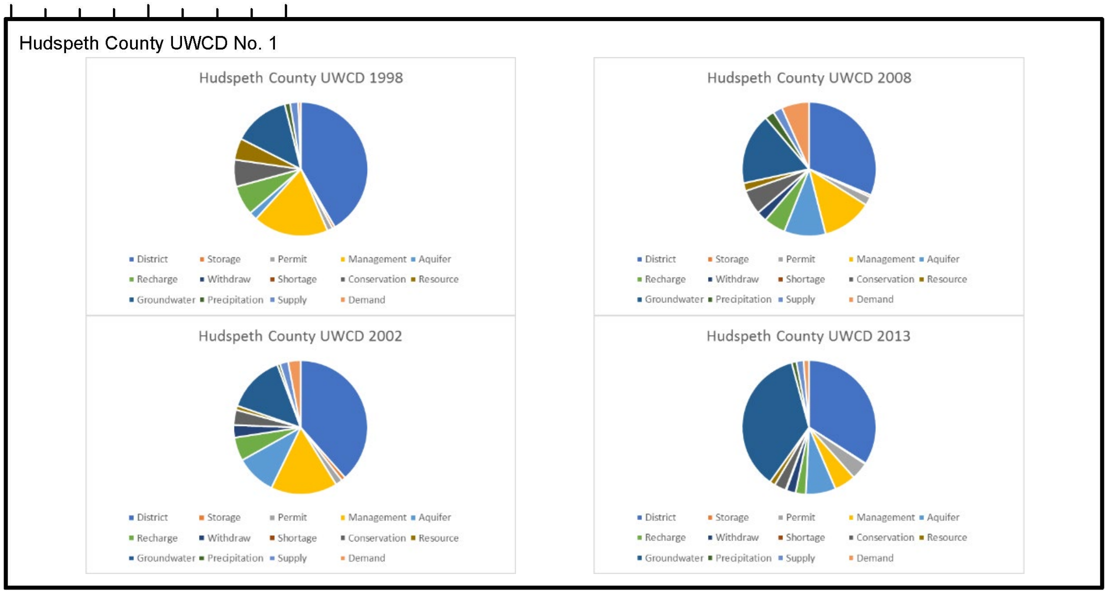


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

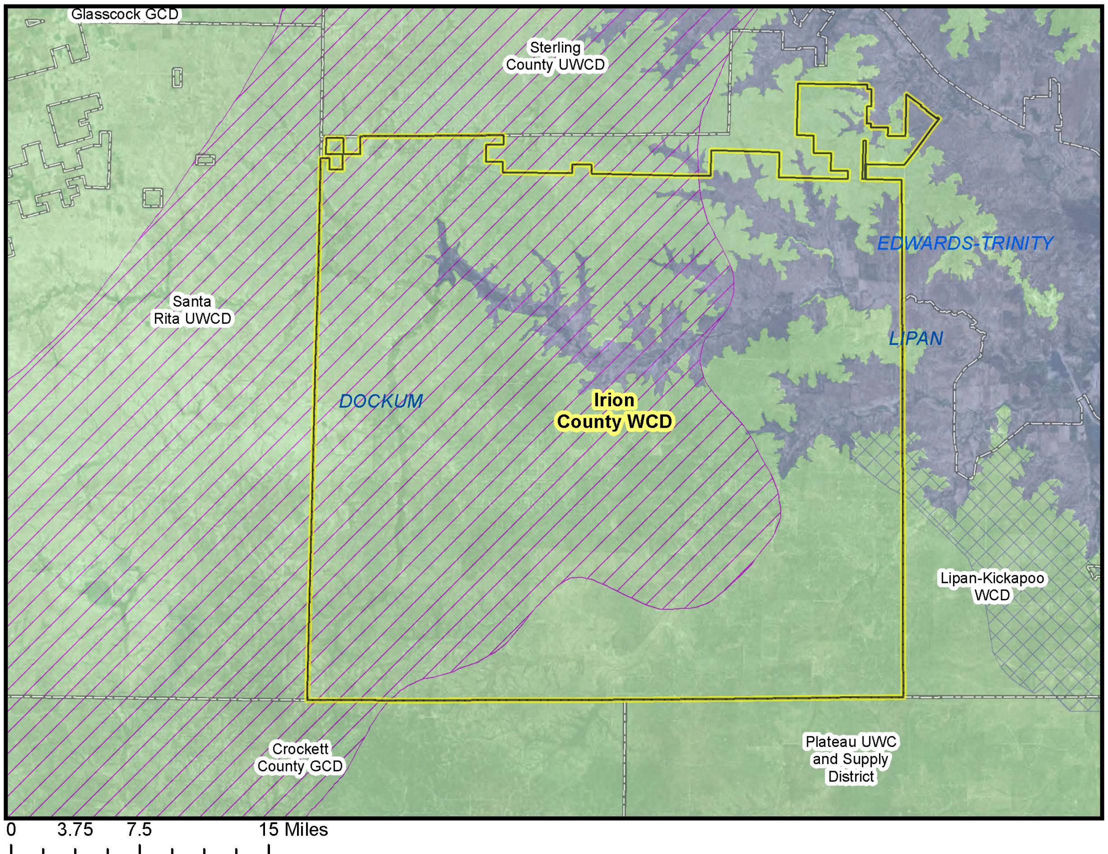


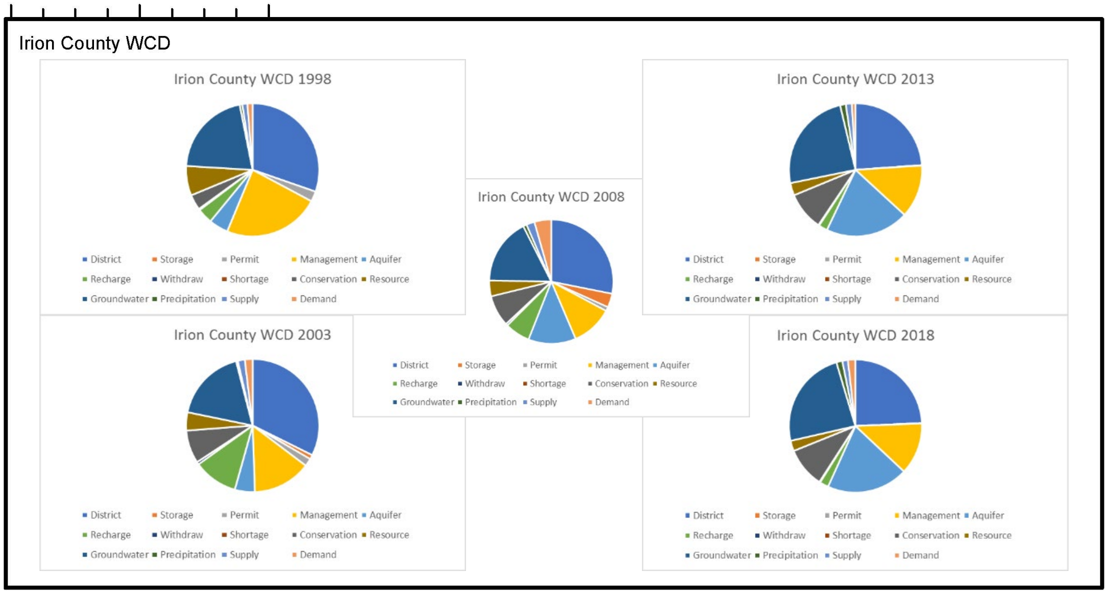
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



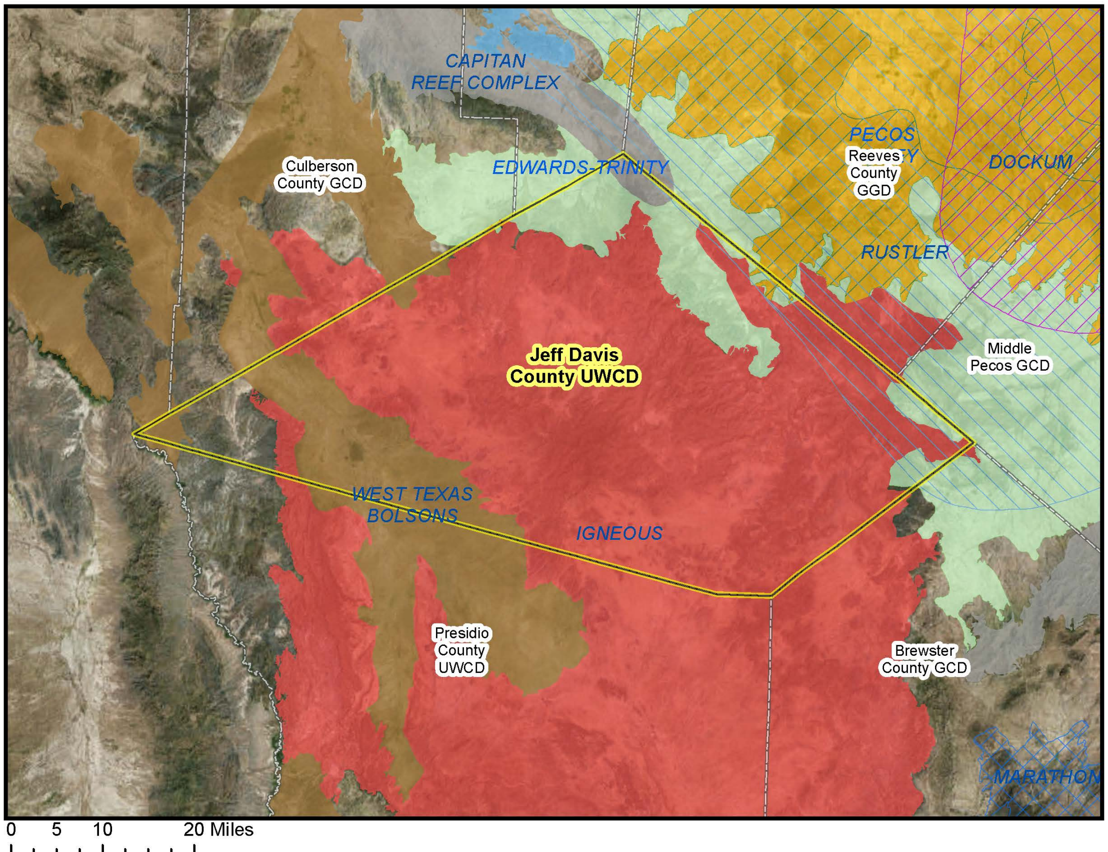


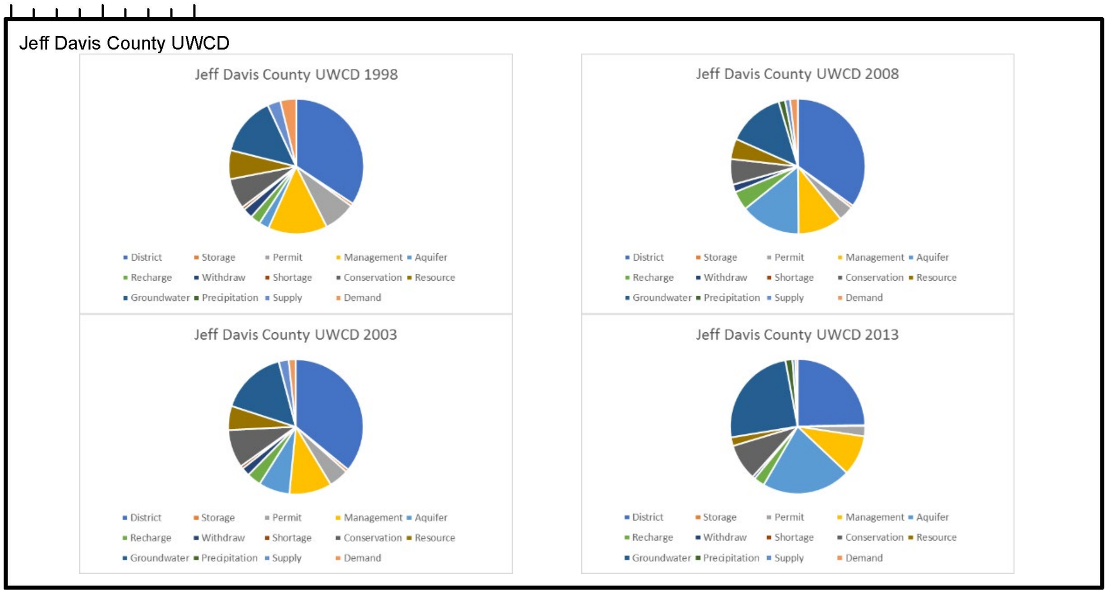
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



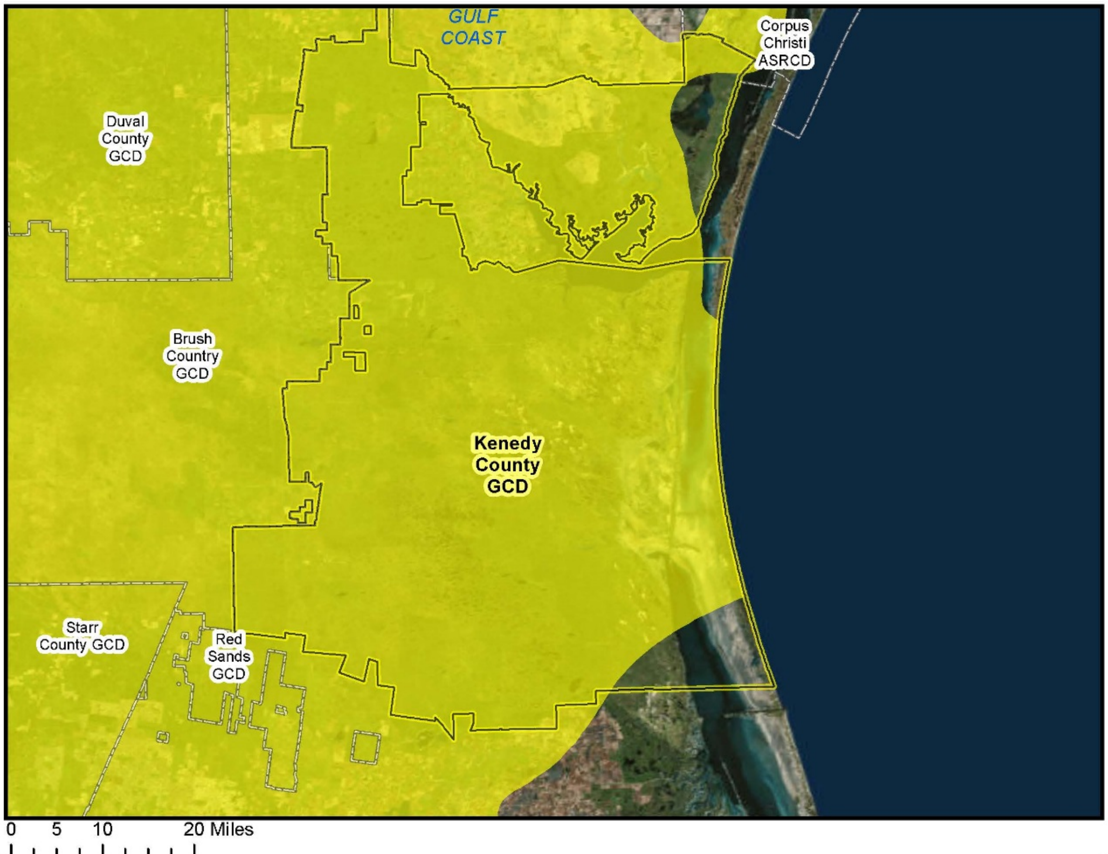


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

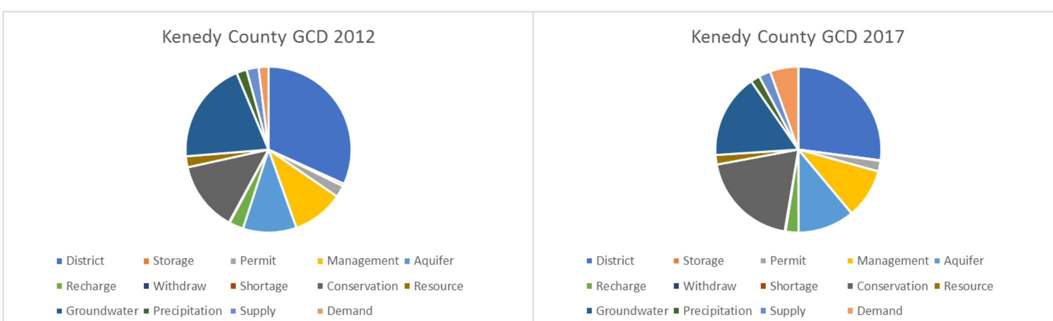




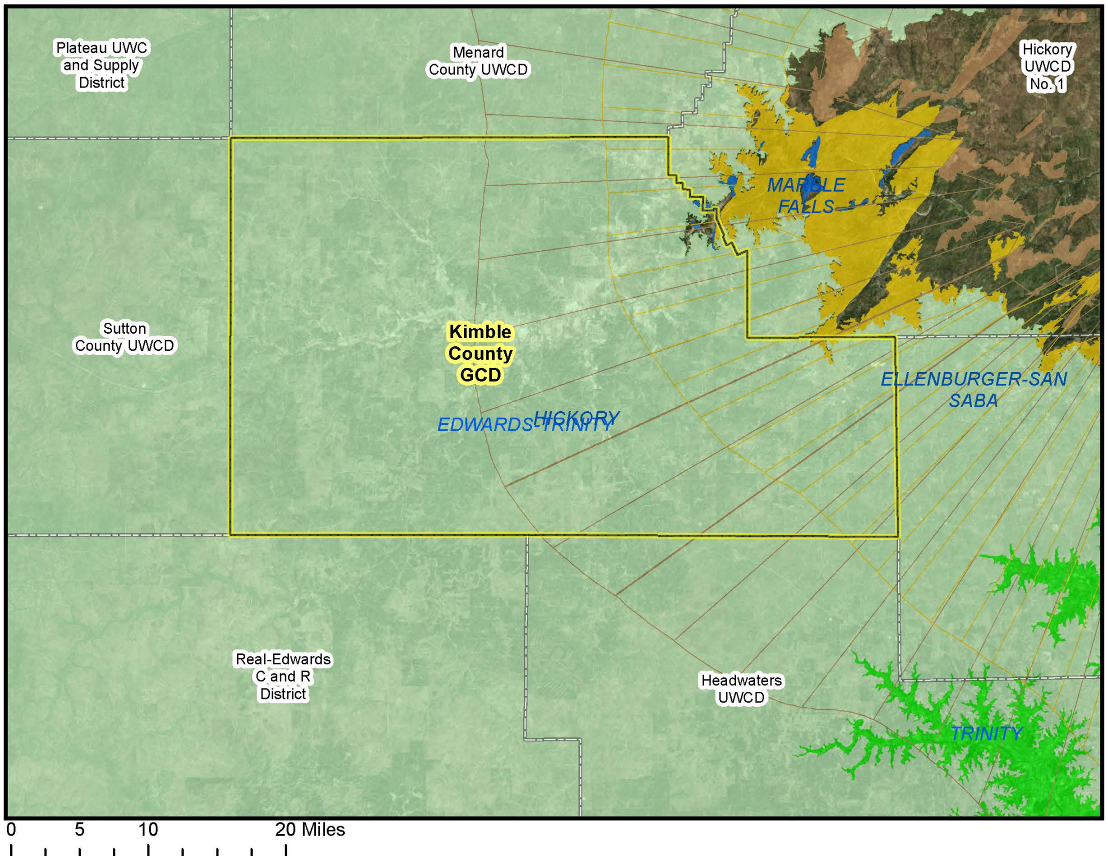
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

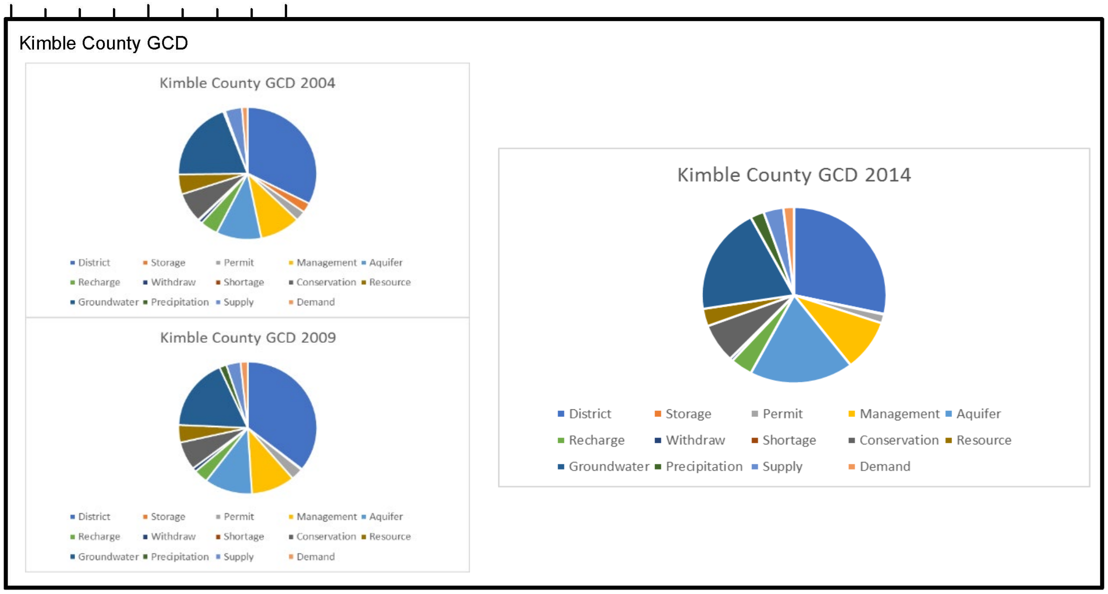


Kenedy County GCD

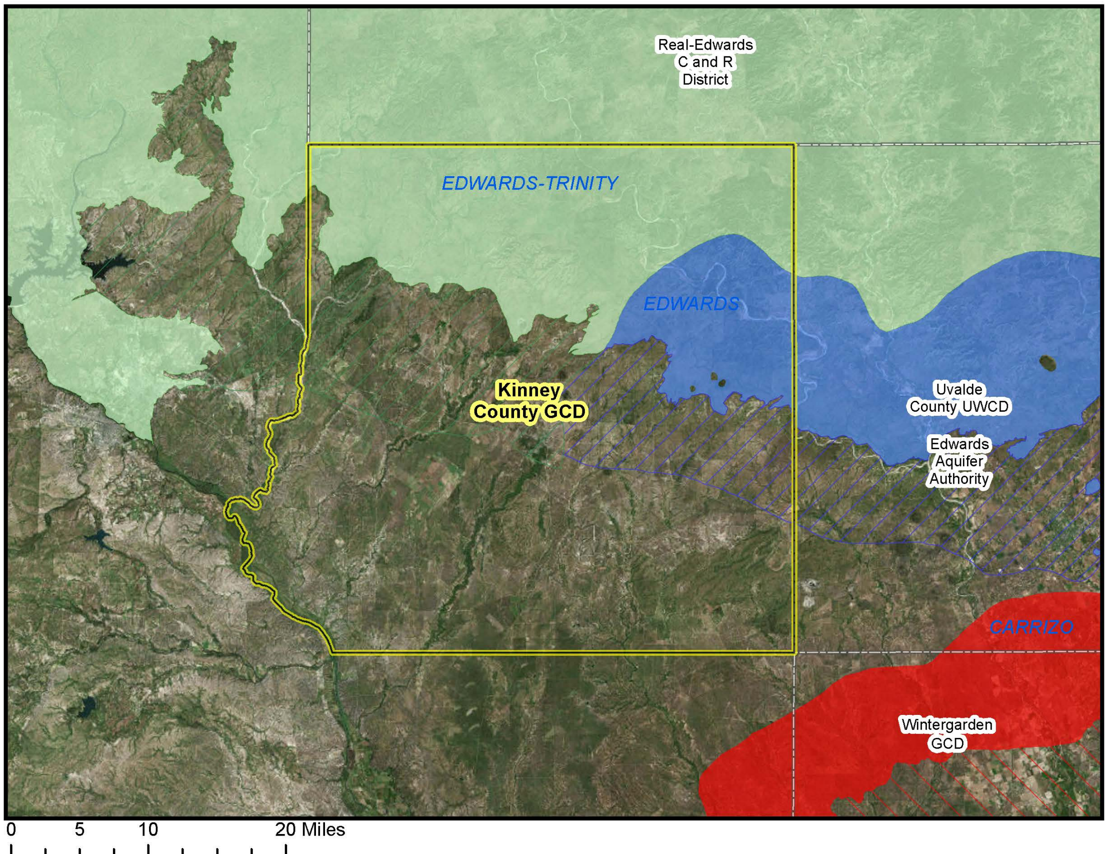


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

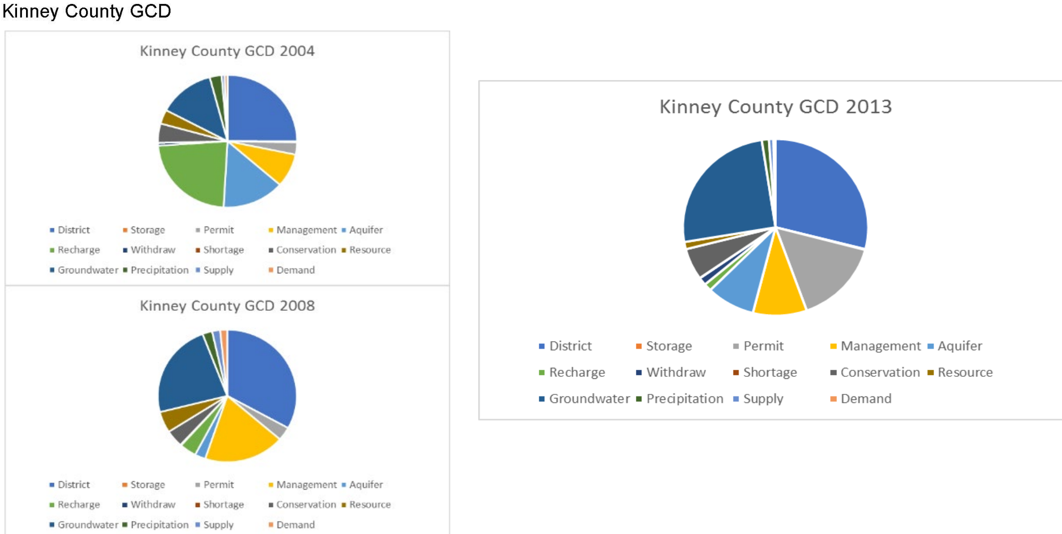




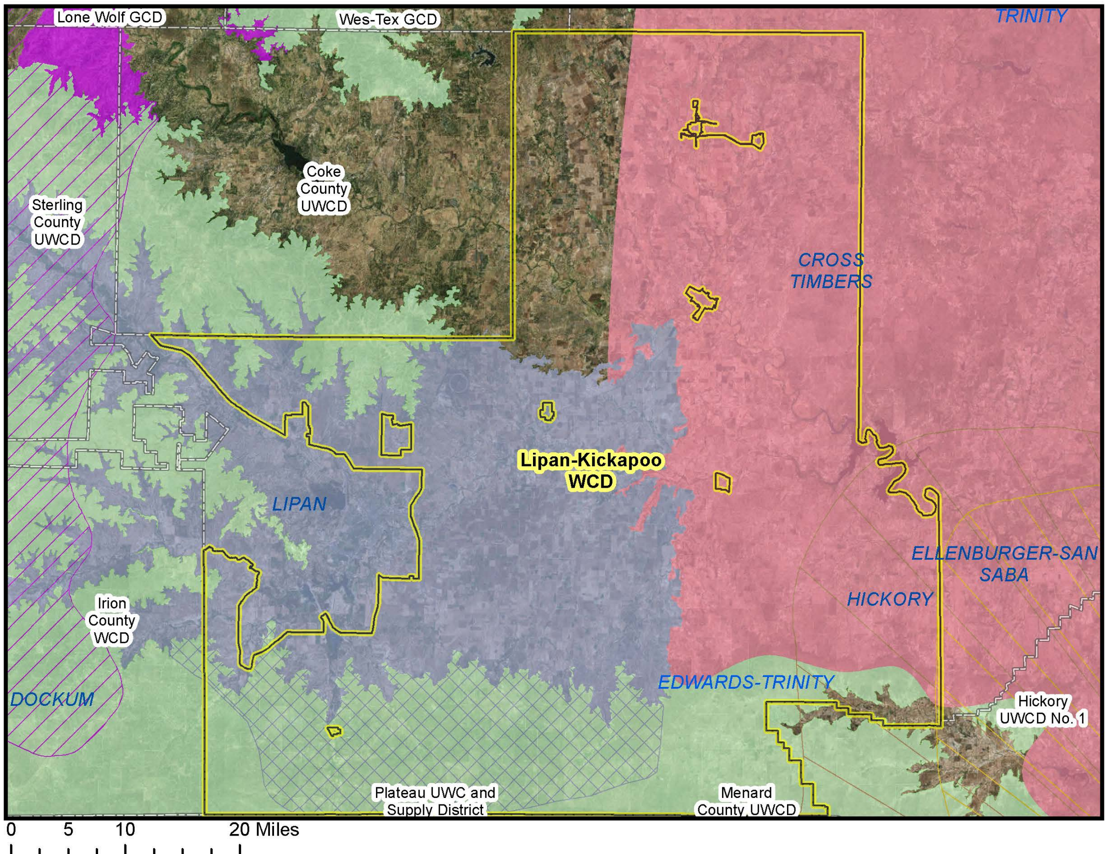
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

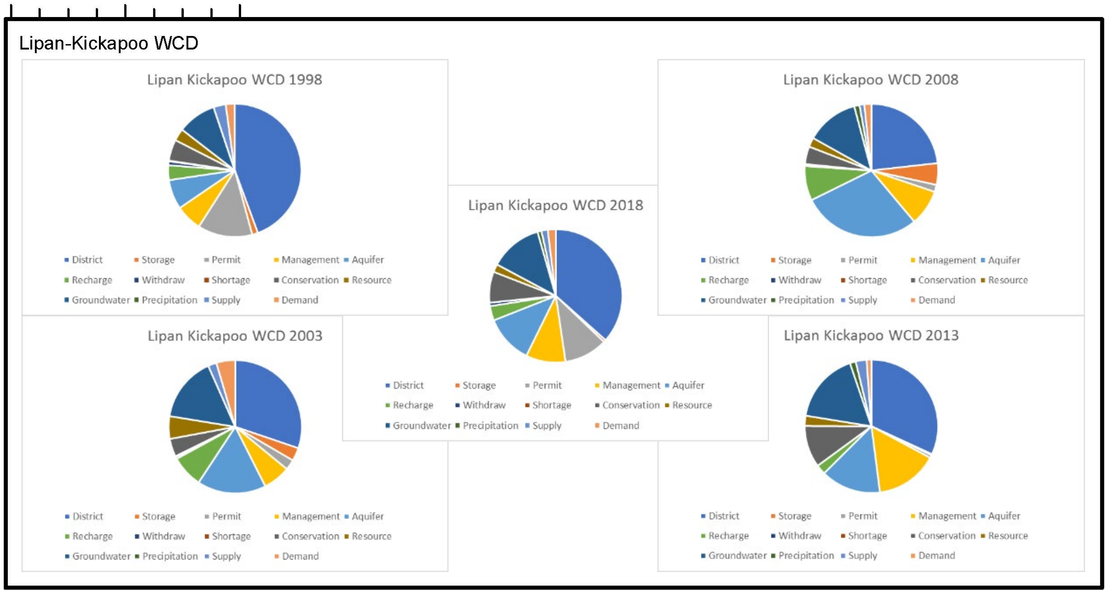


Kinney County GCD

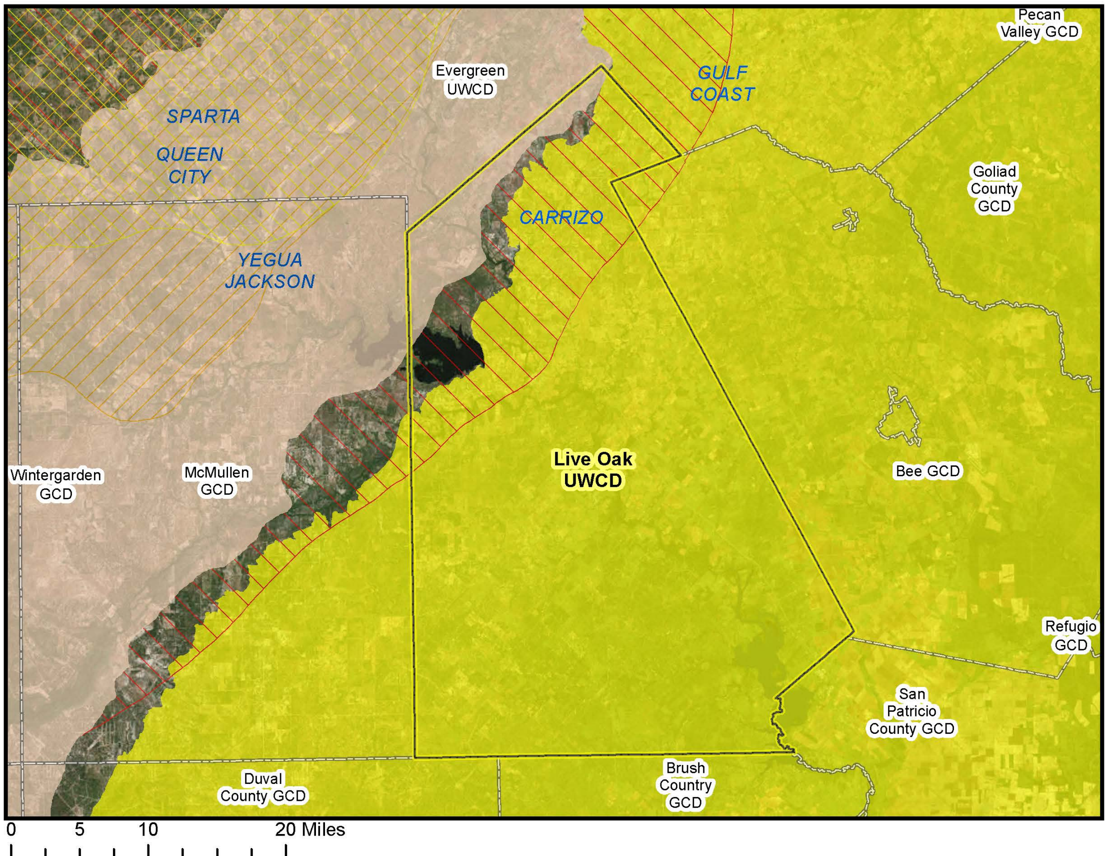


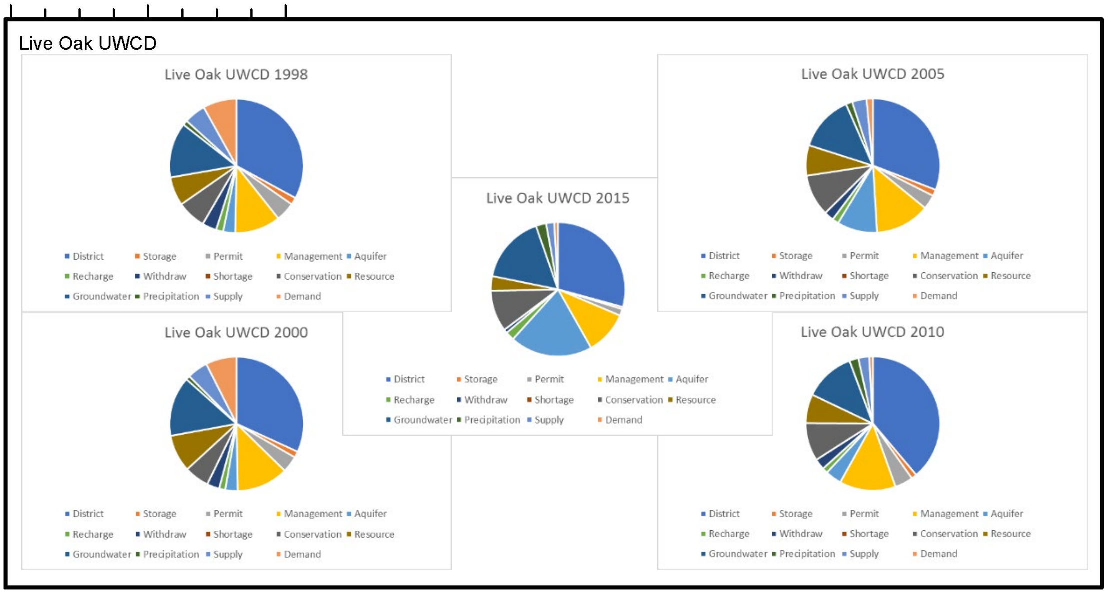
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



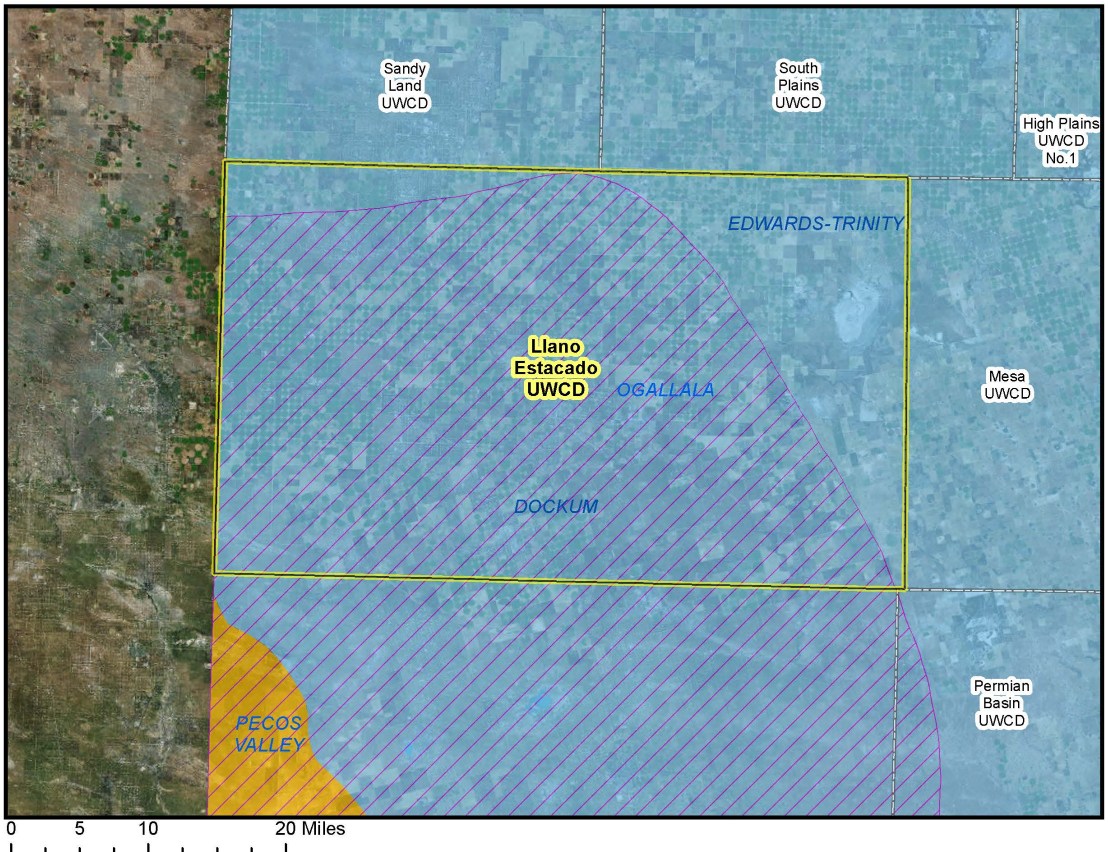


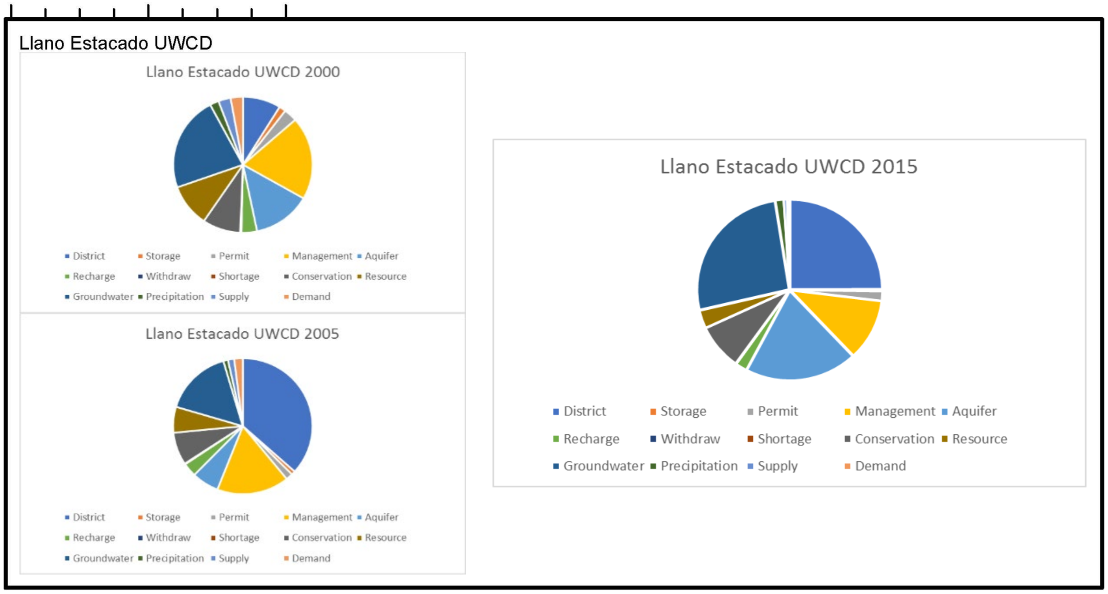
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



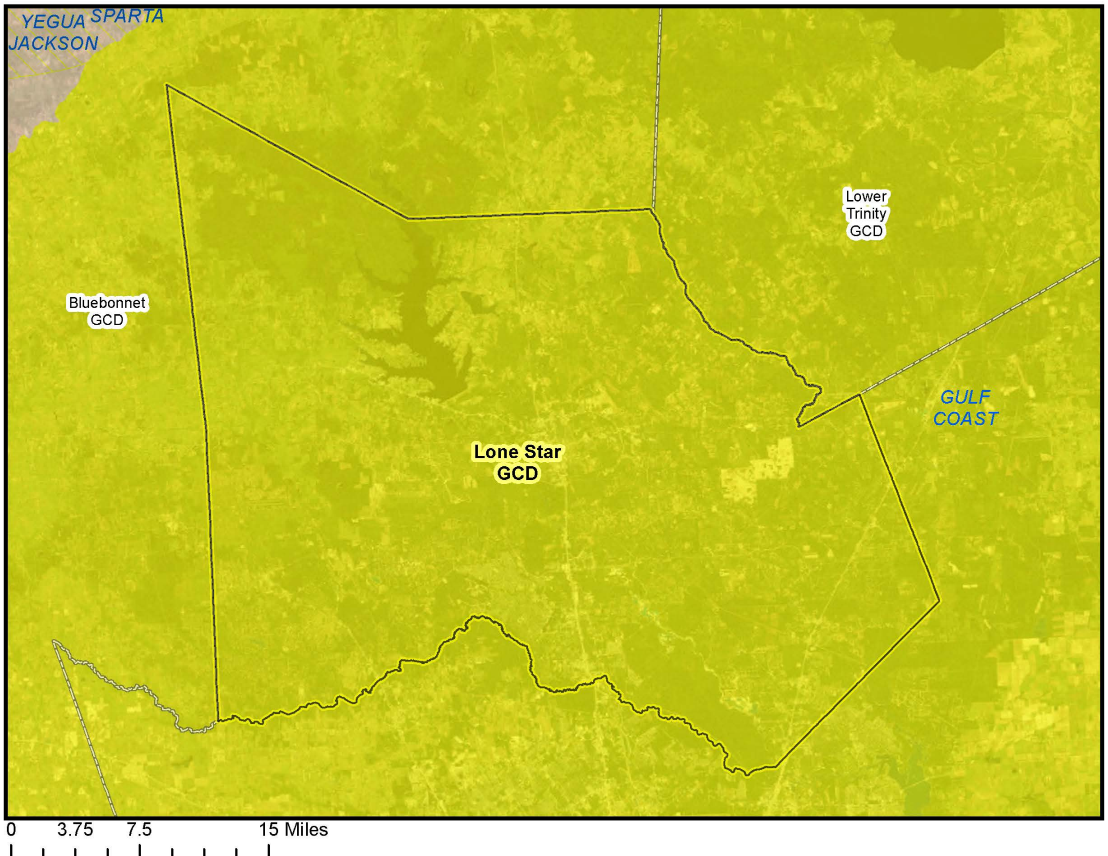


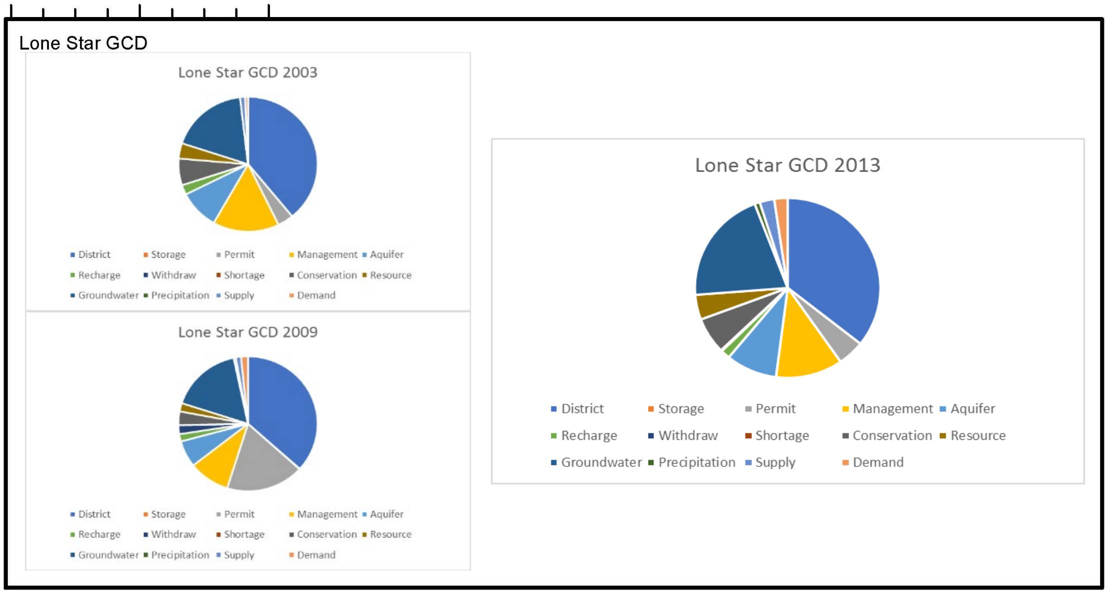
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



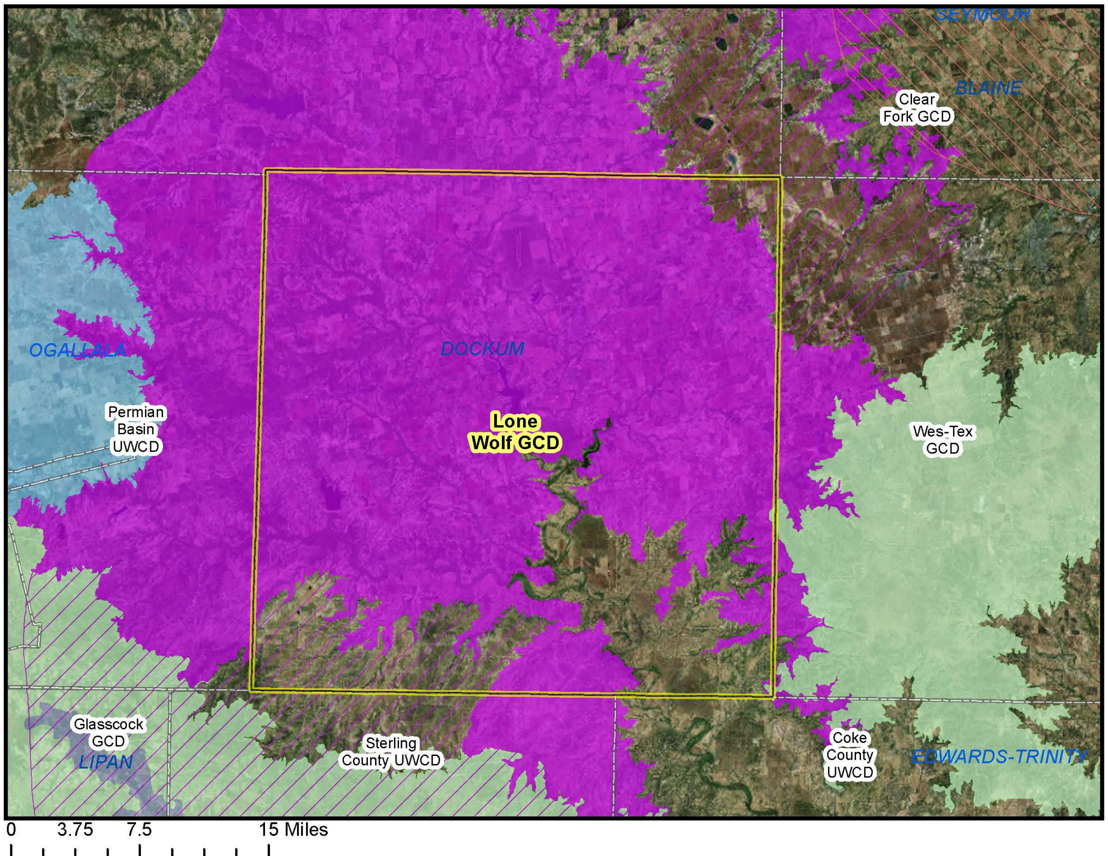


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

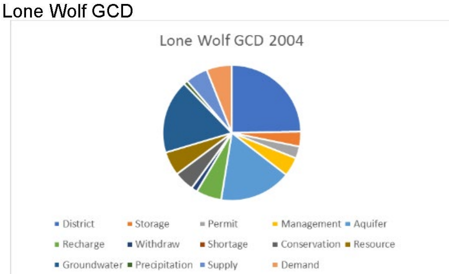


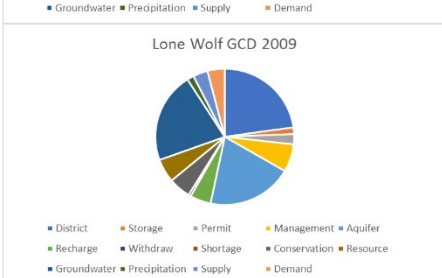


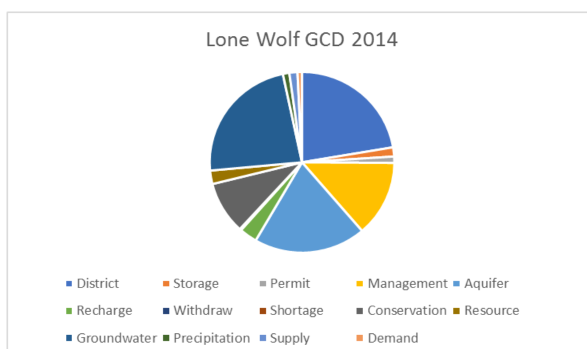
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



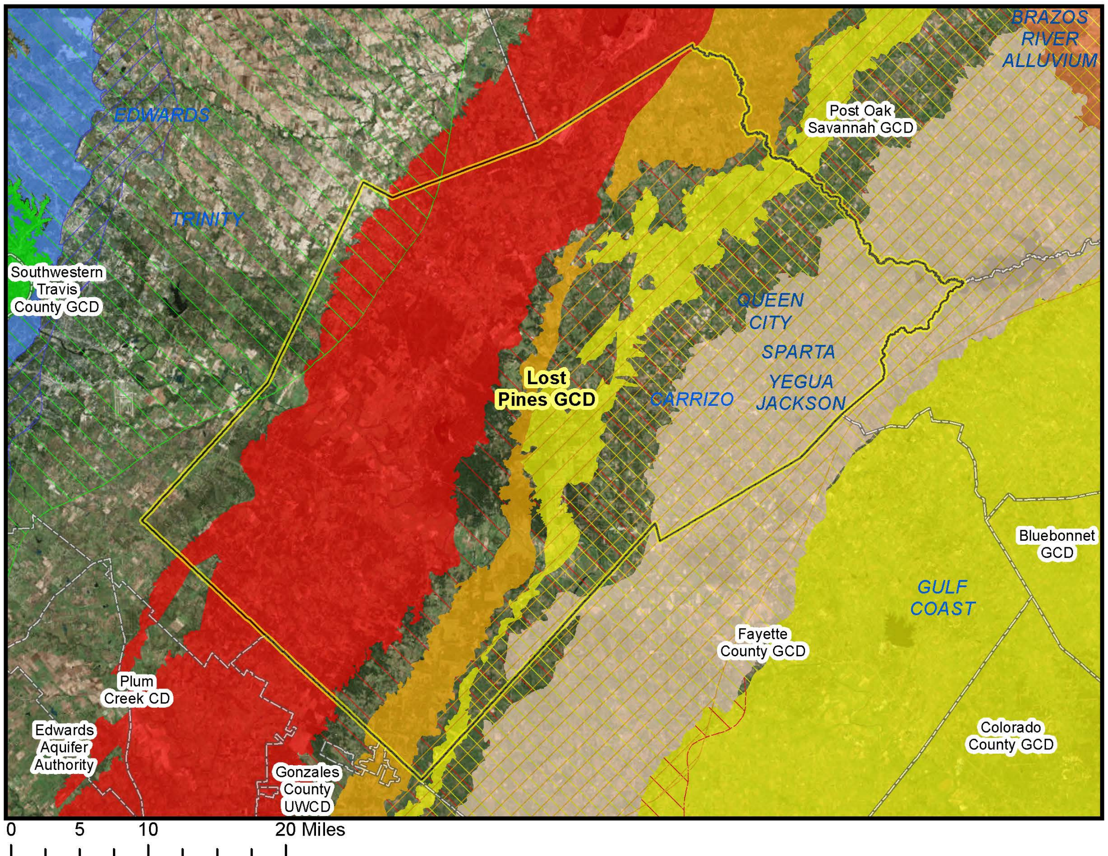
Lone Wolf GCD

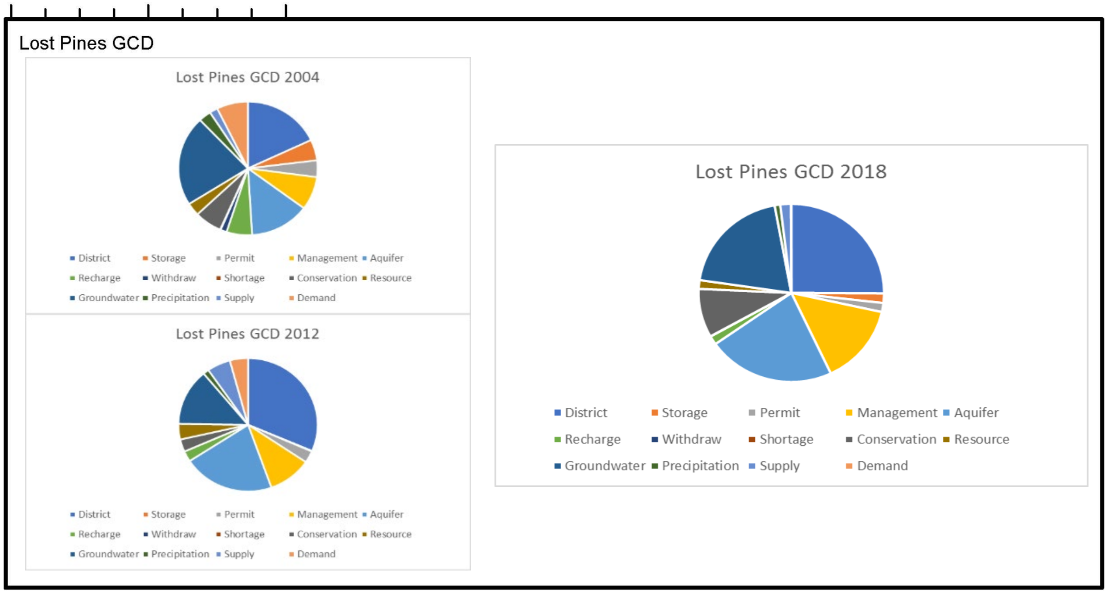




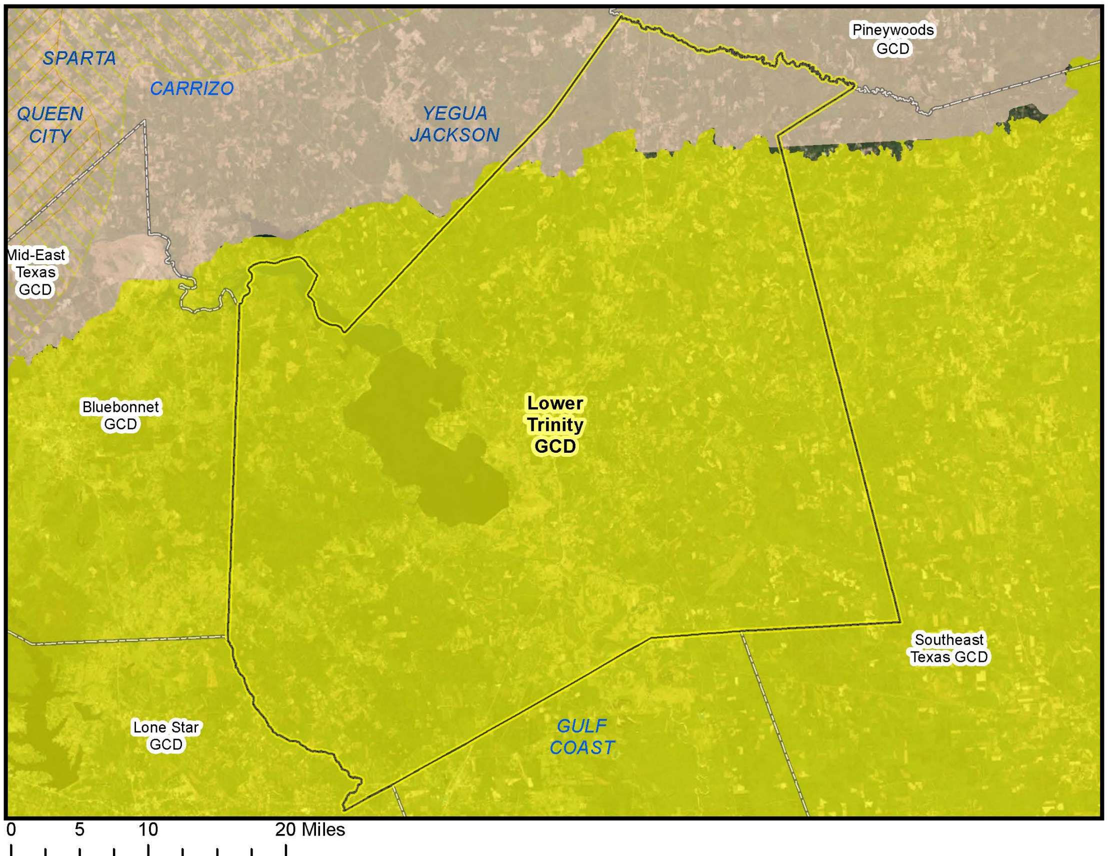


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

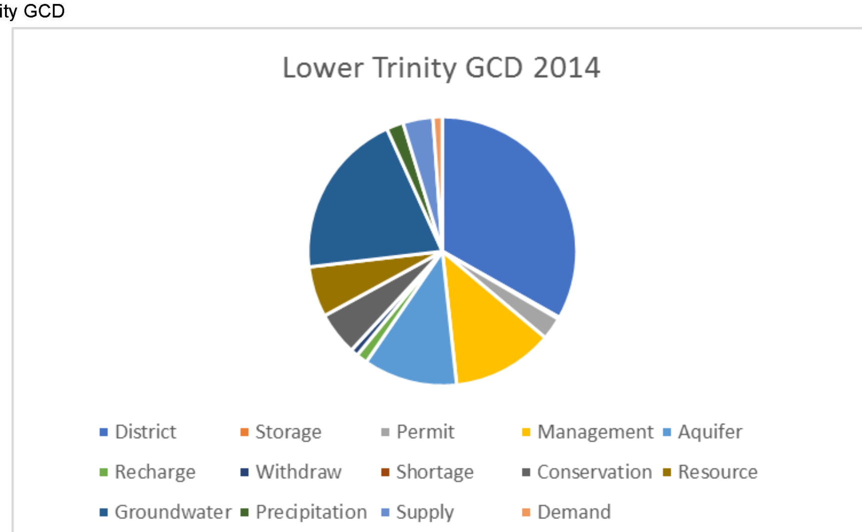




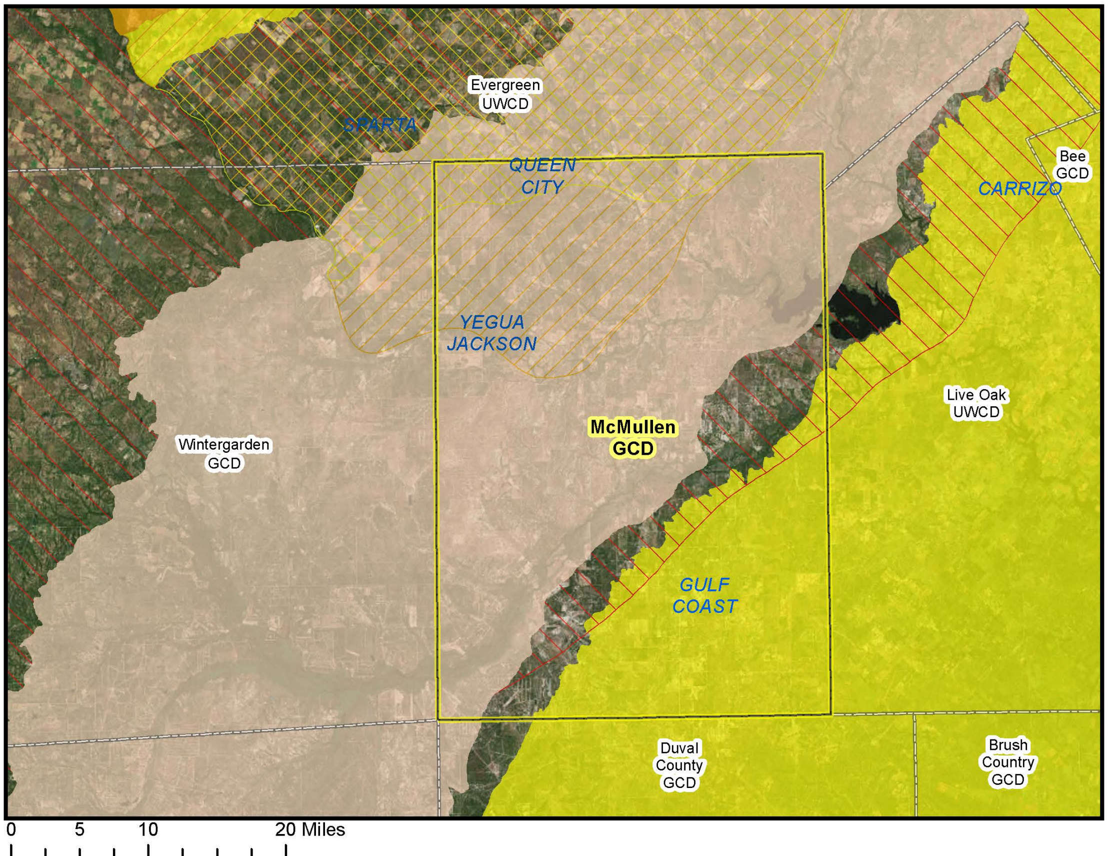
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

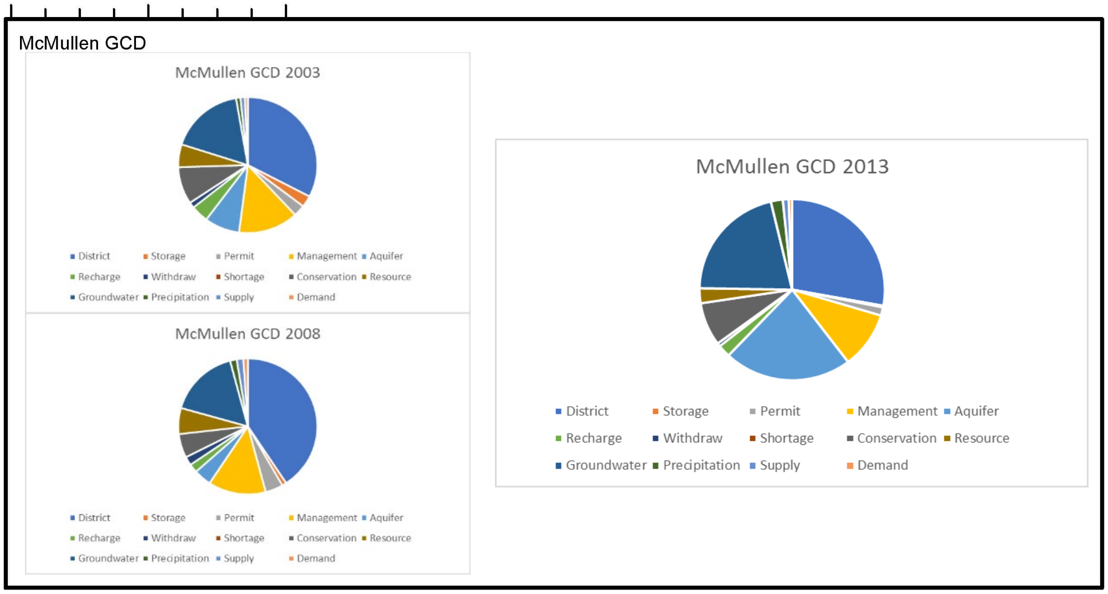


Lower Trinity GCD

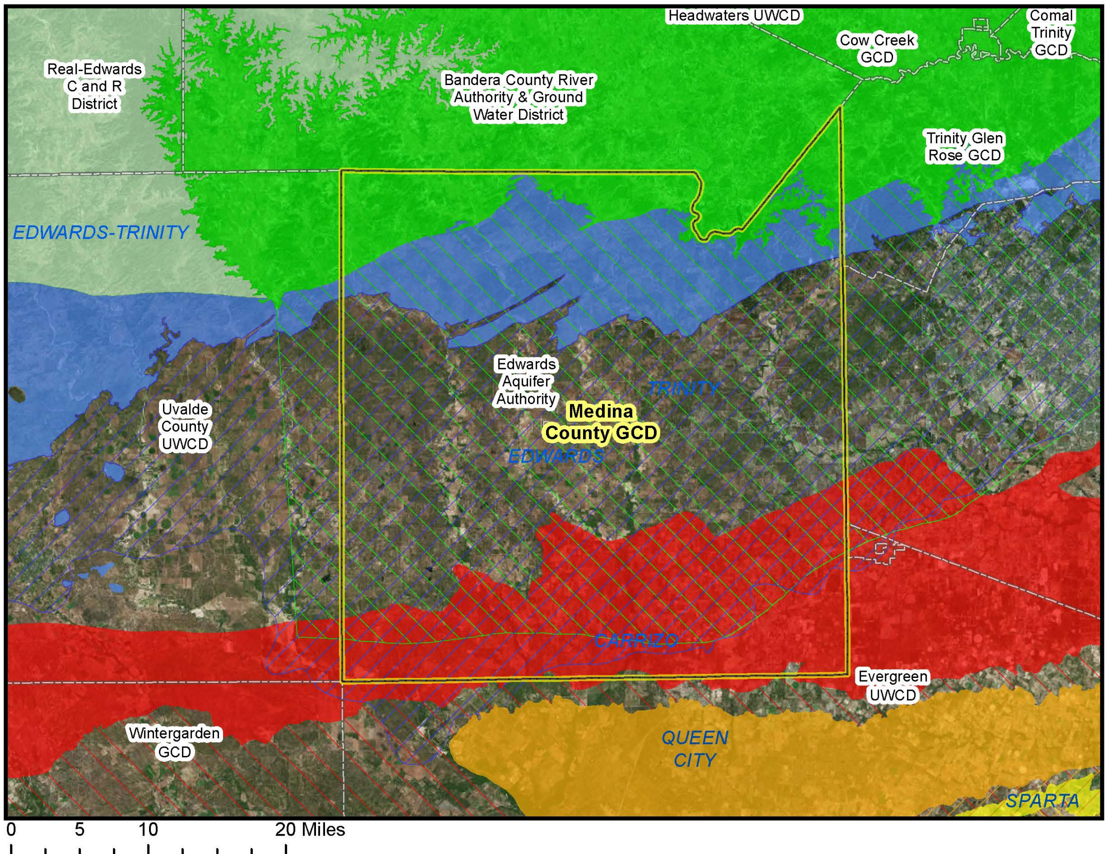


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

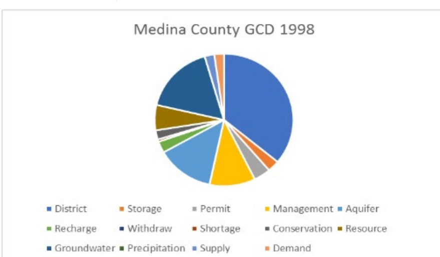


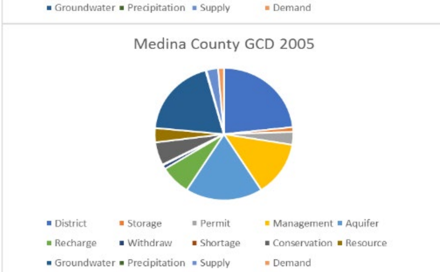


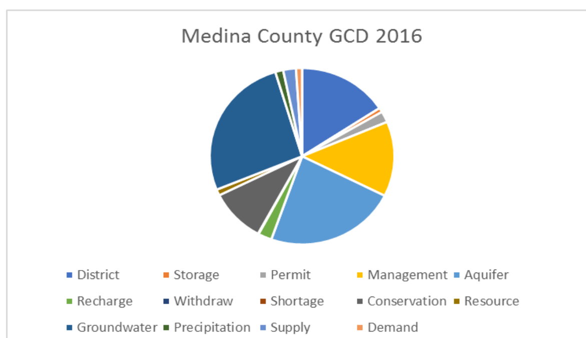
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



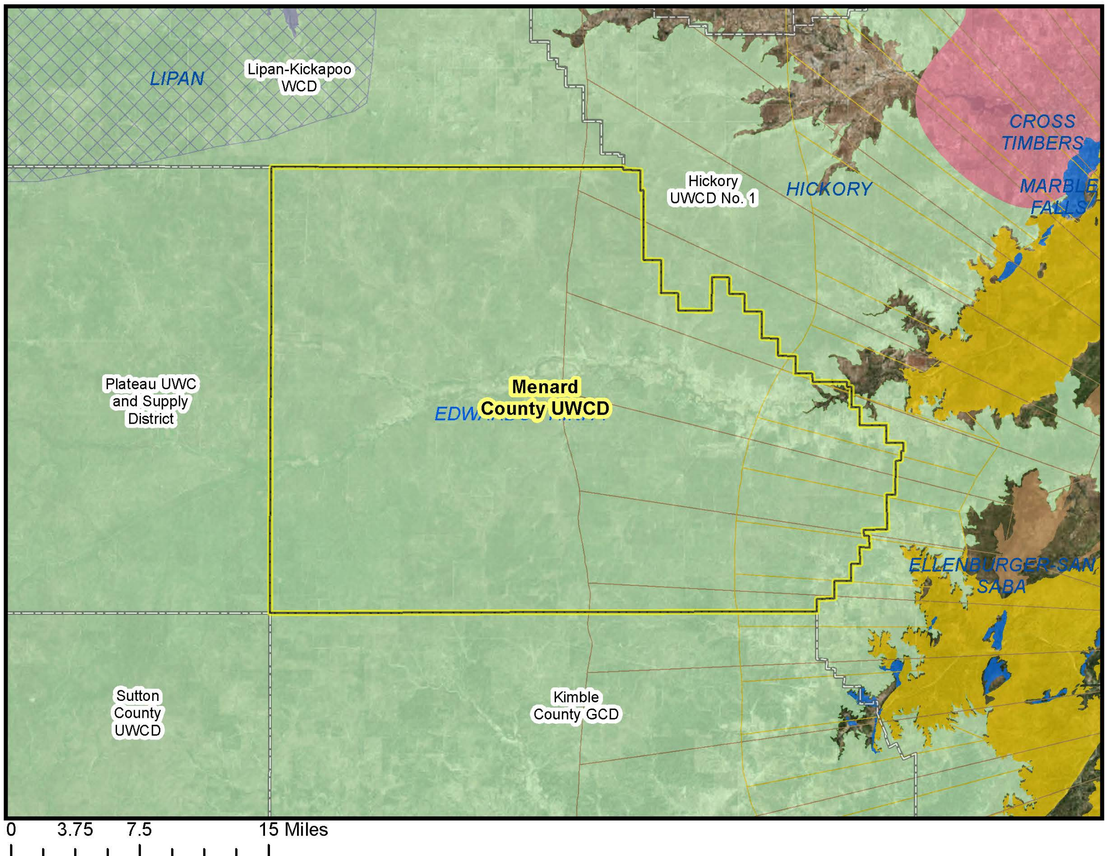
Medina County GCD

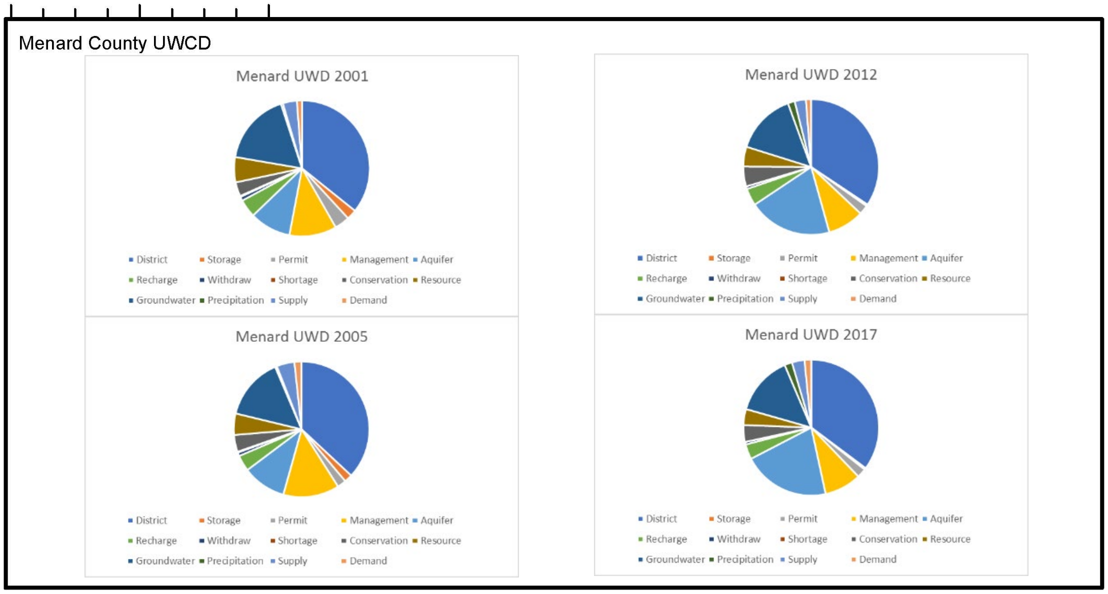




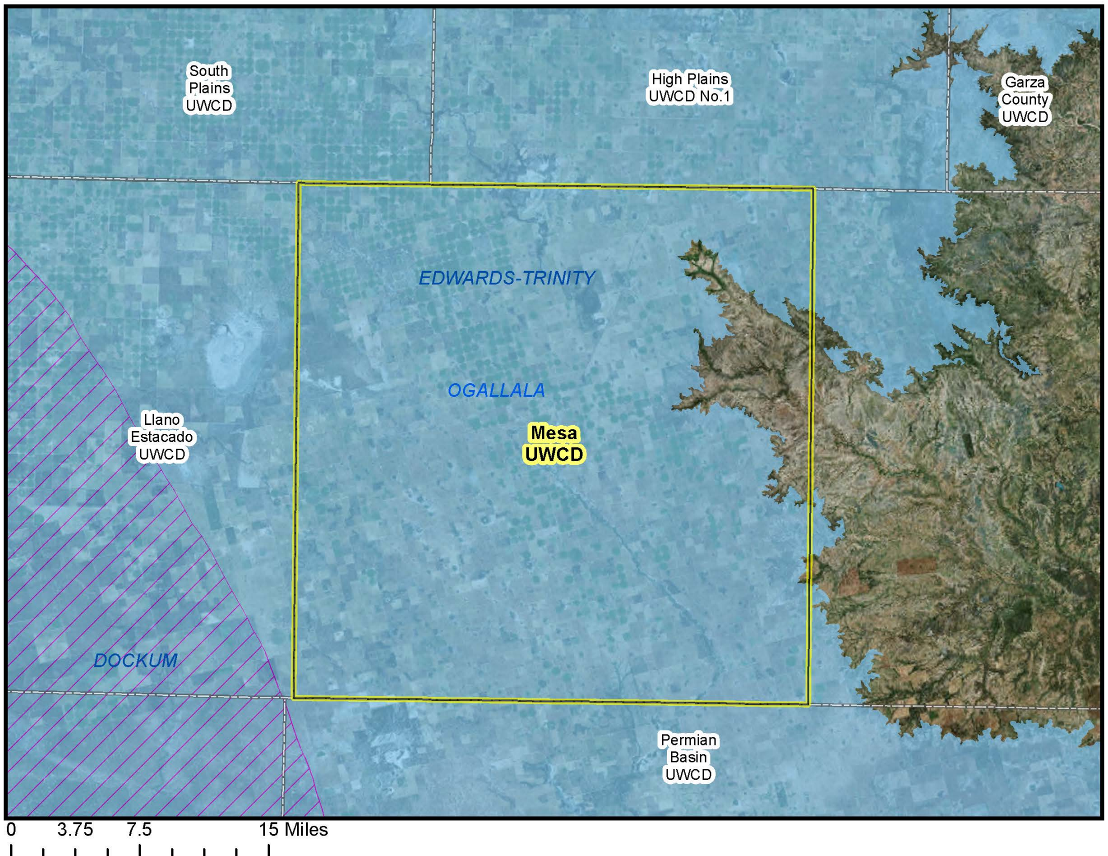


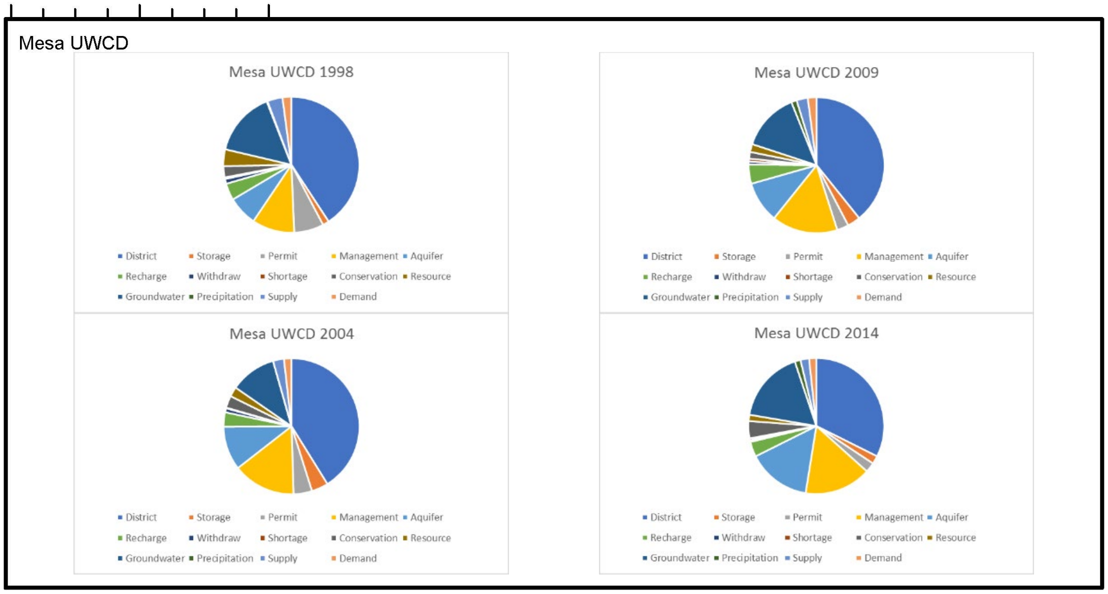
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



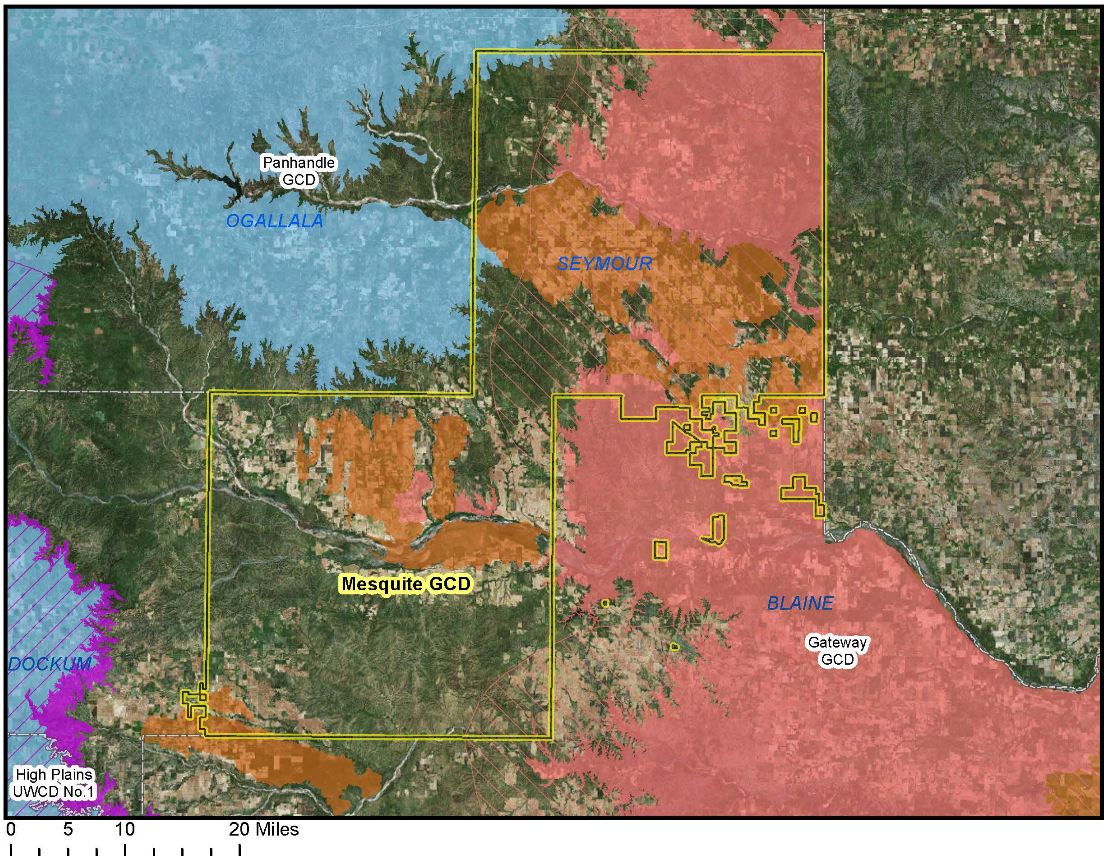


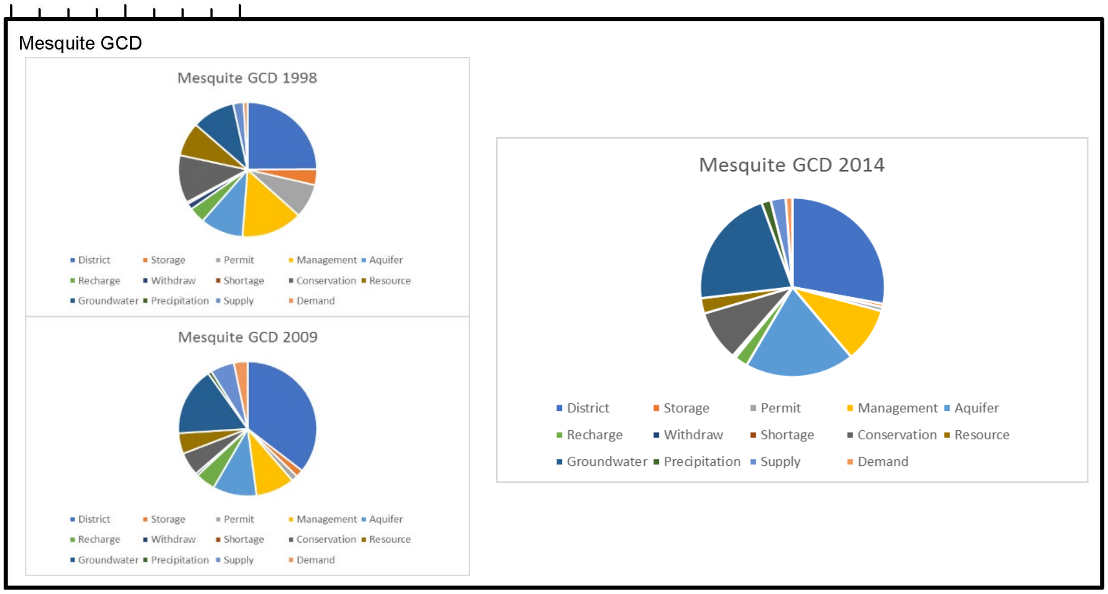
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



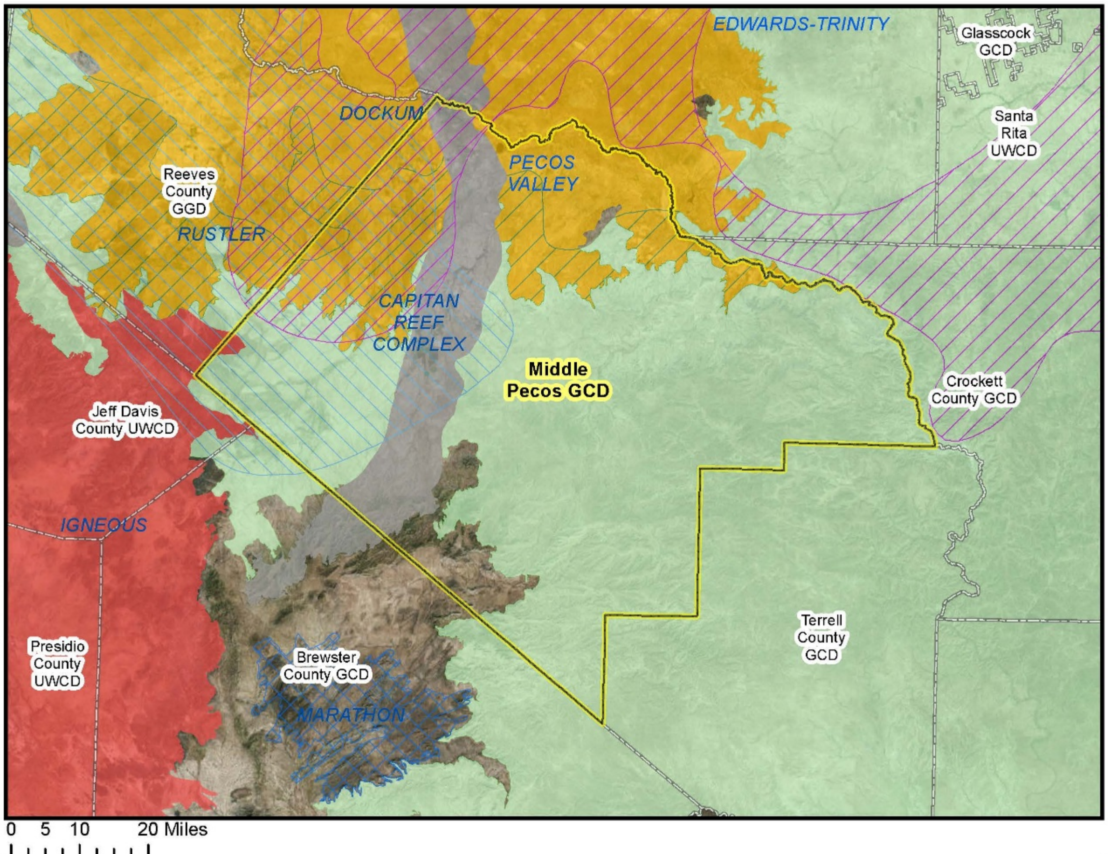


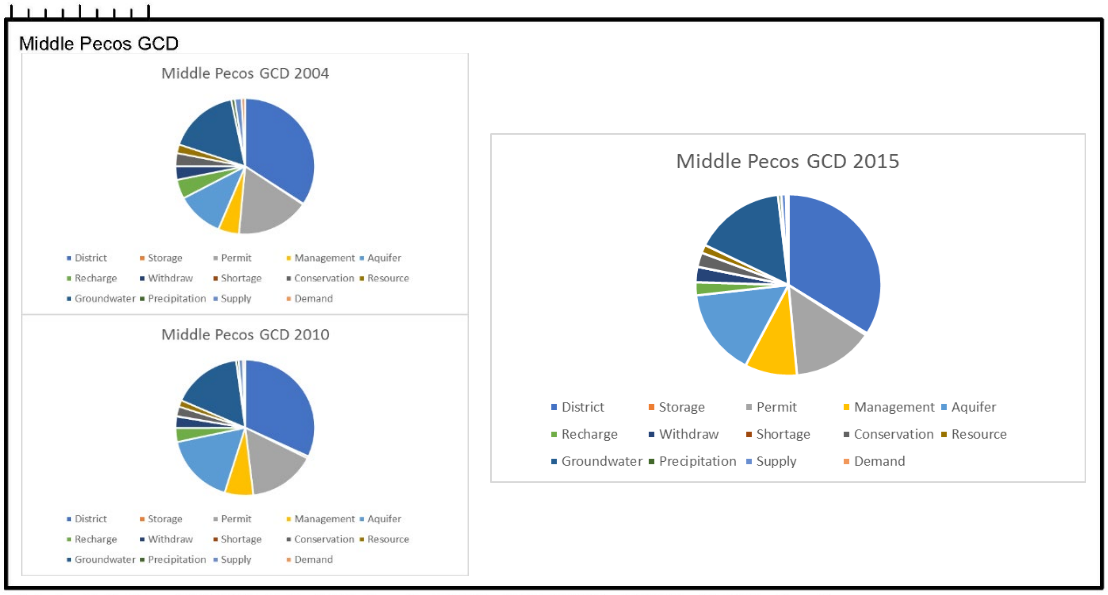
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



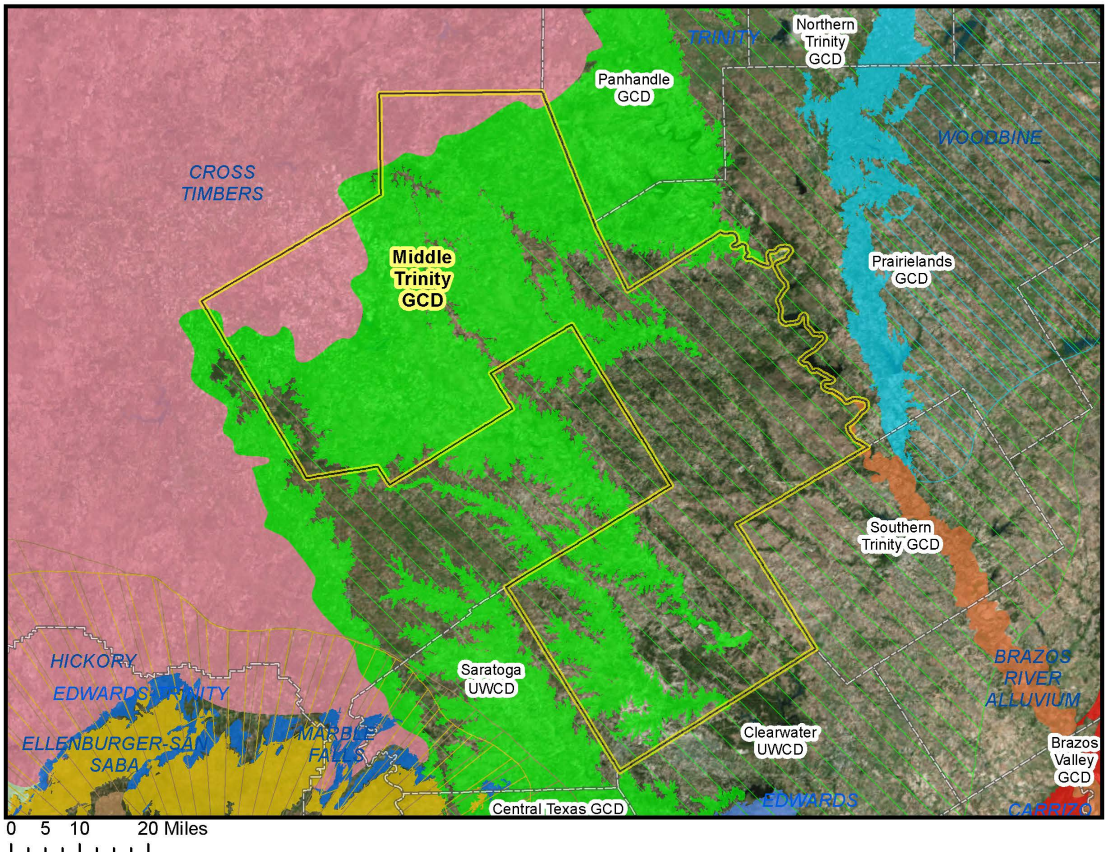


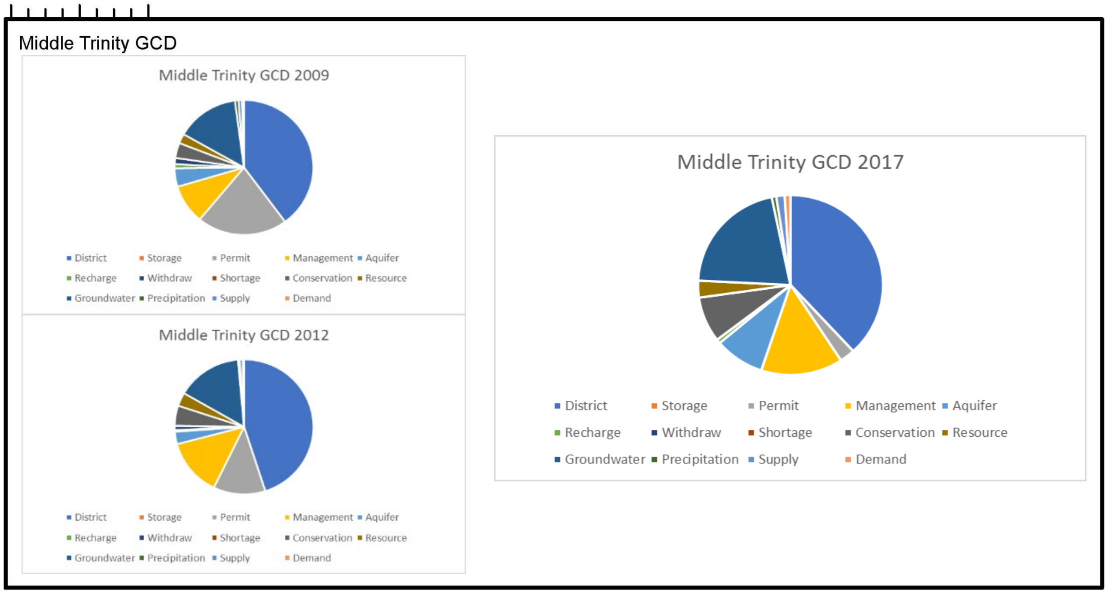
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



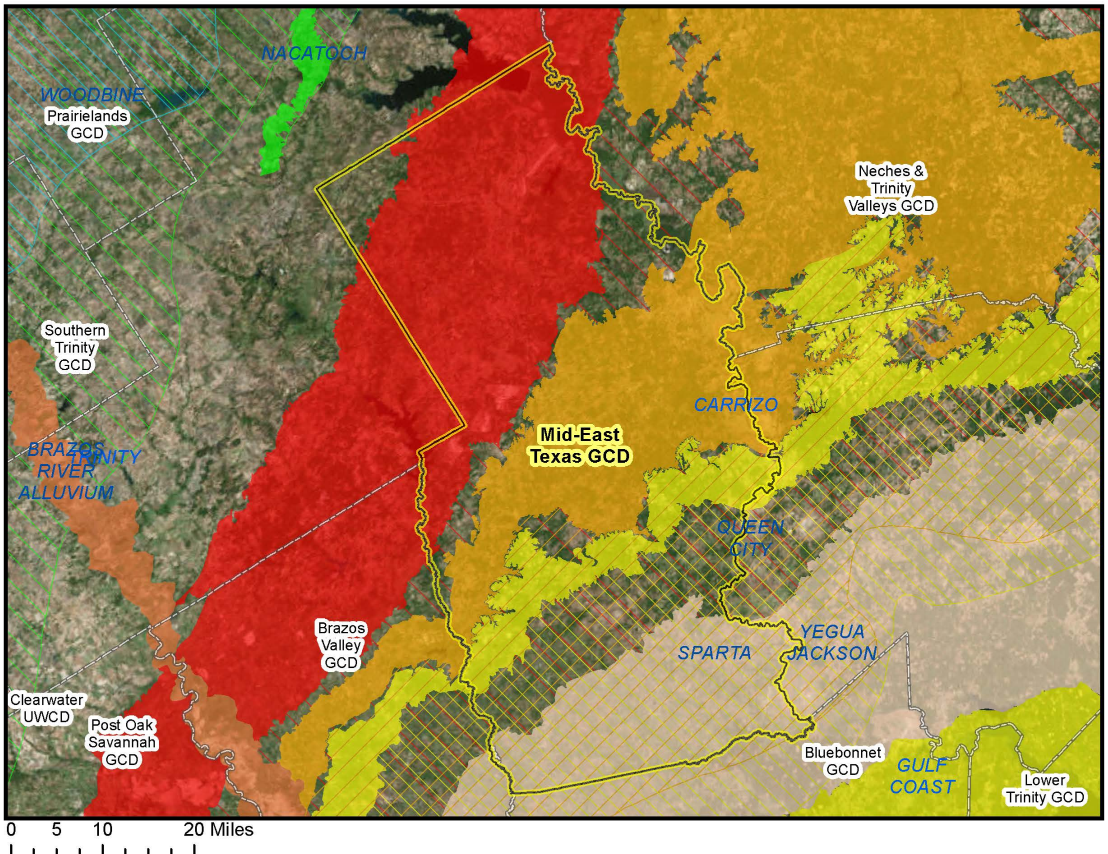


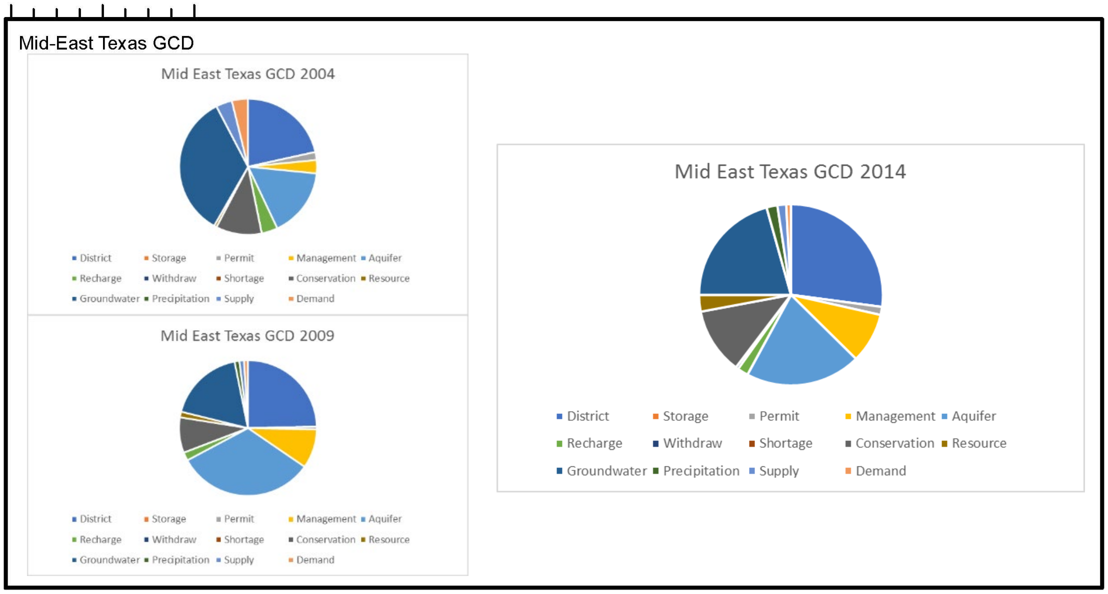
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



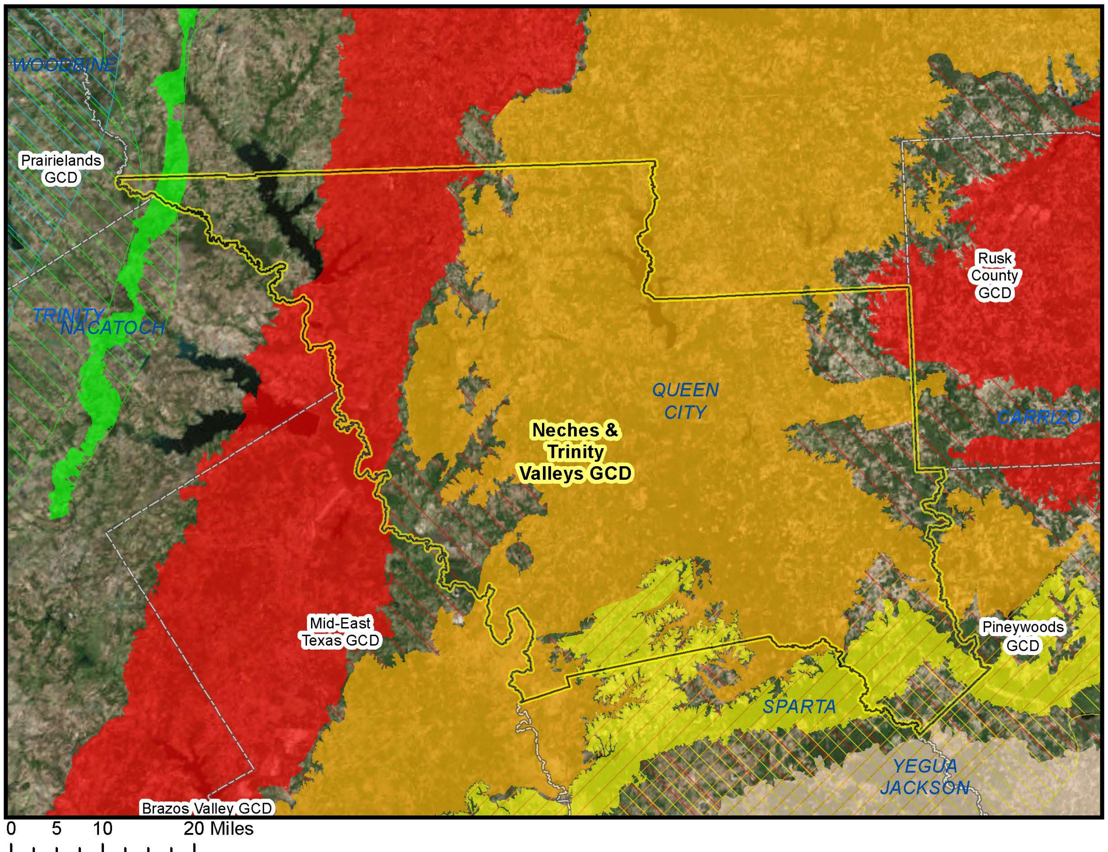


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

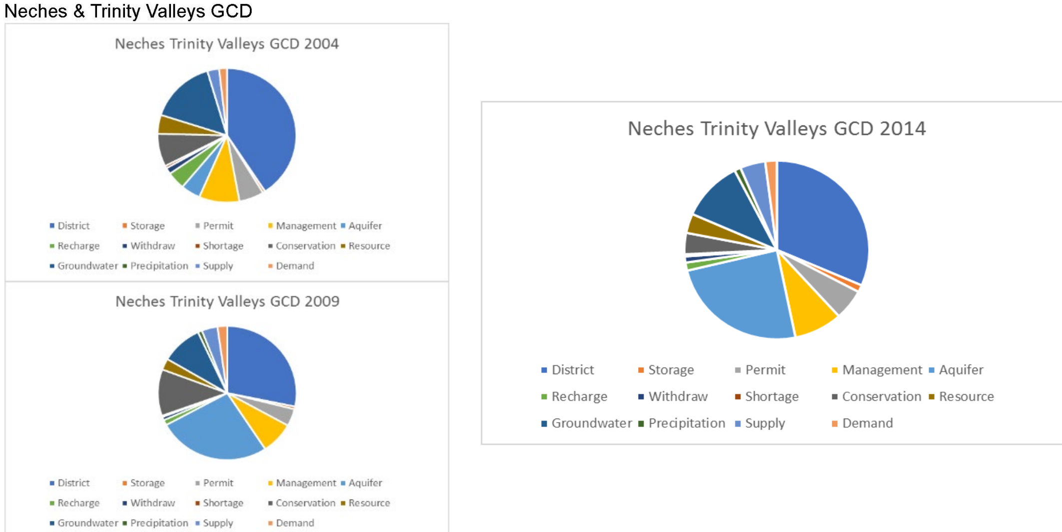




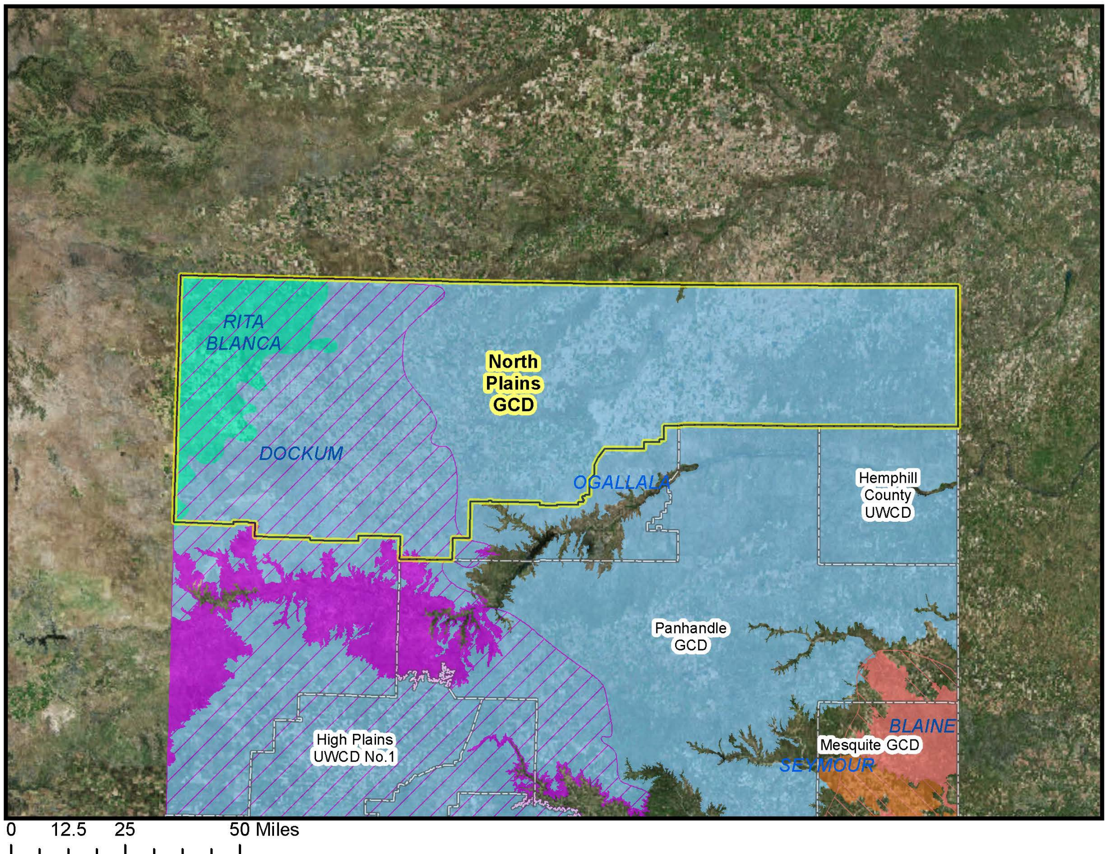
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

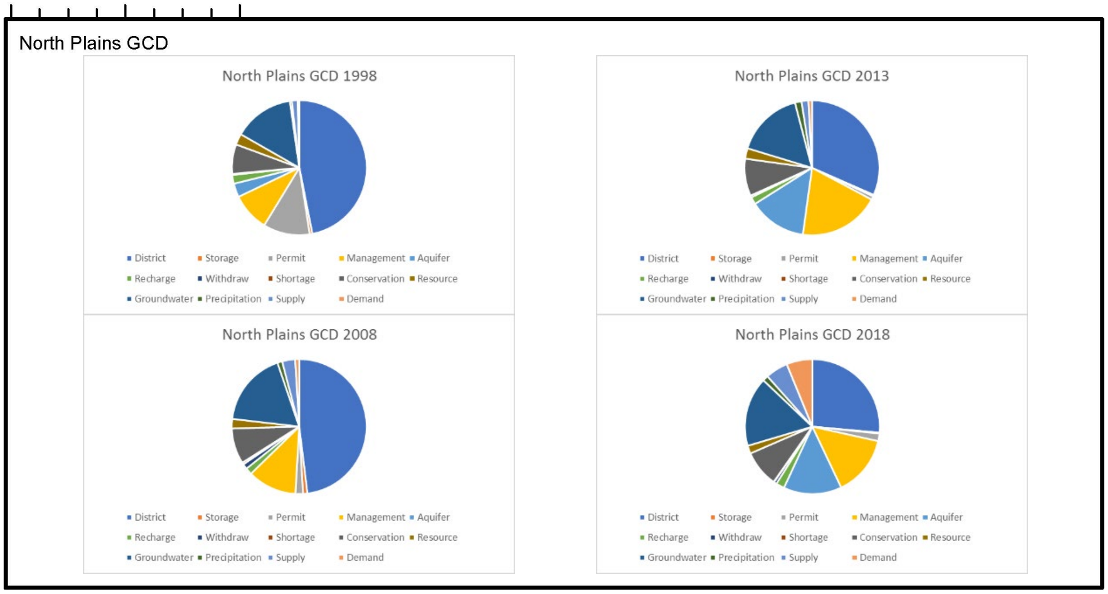


Neches & Trinity Valleys GCD

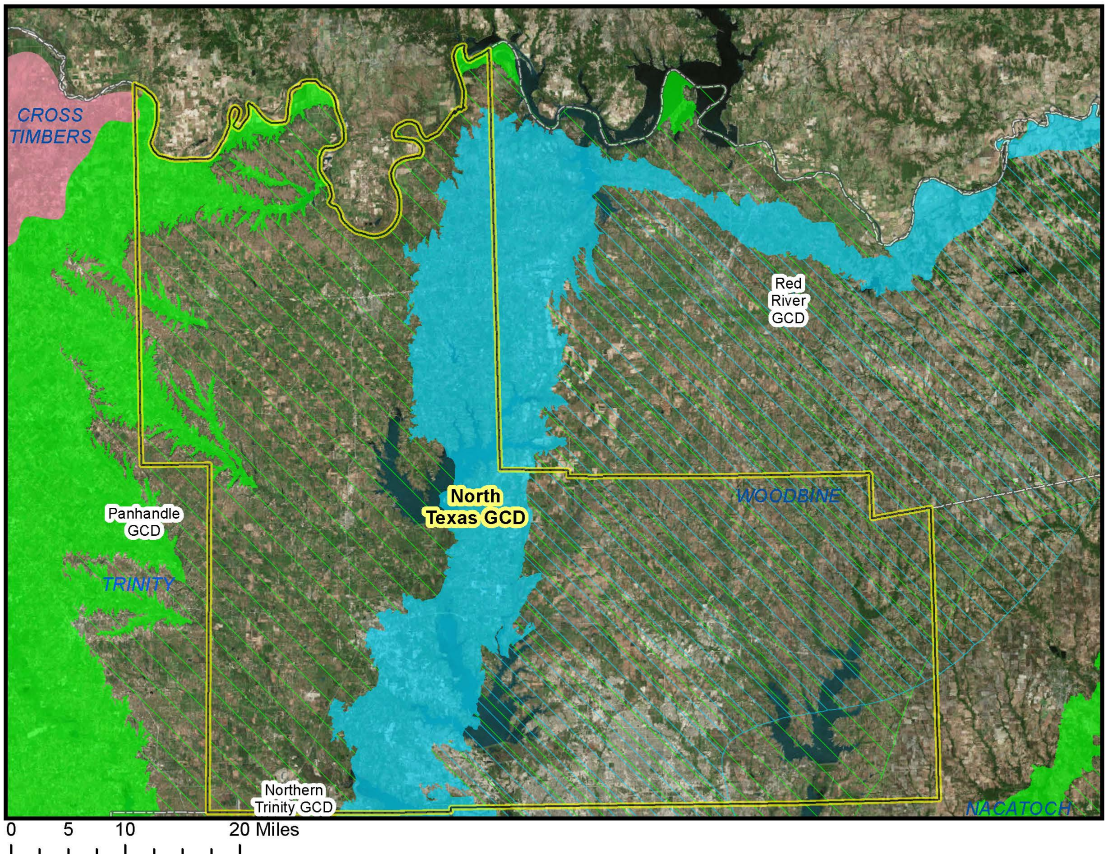


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

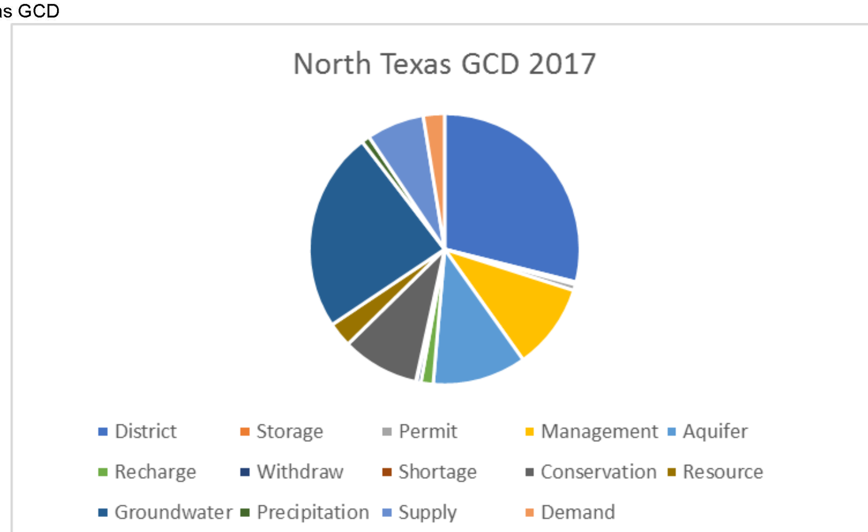




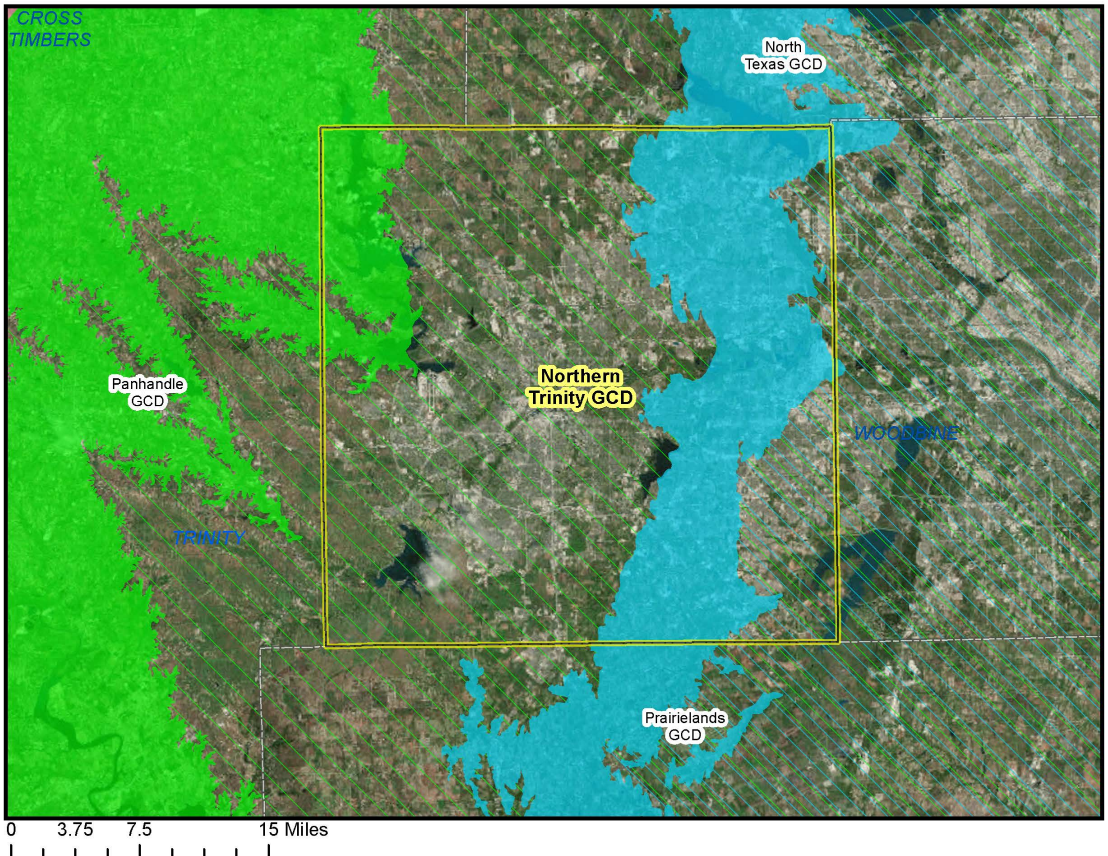
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



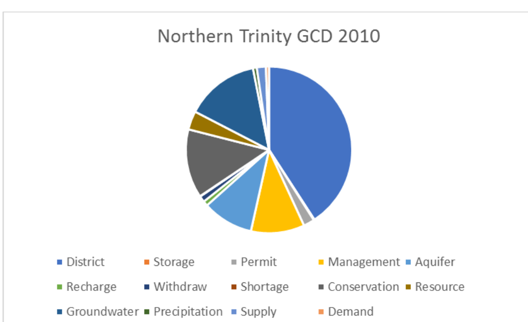
North Texas GCD

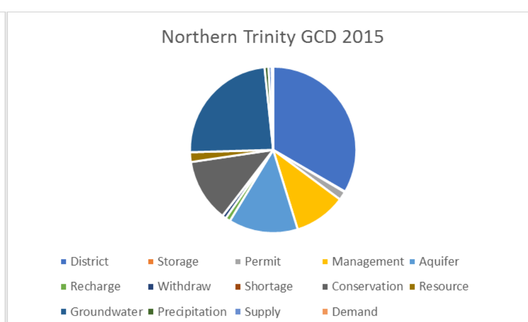


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

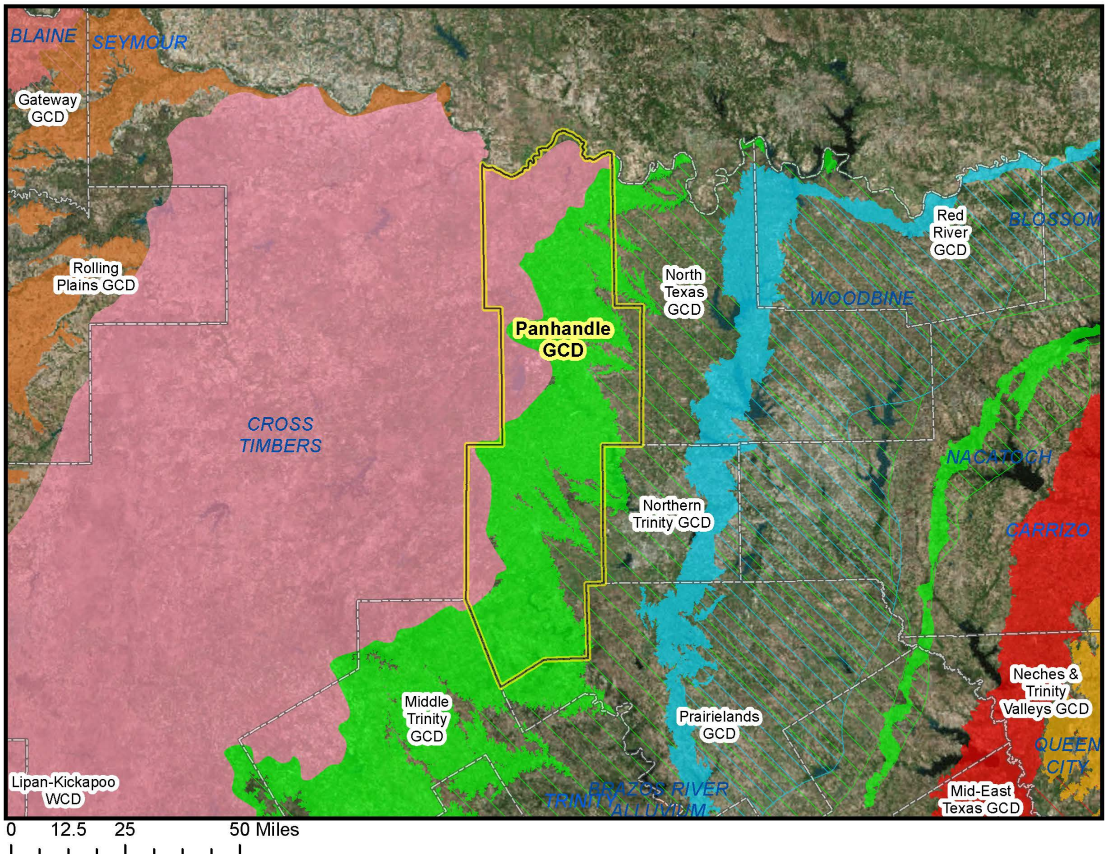


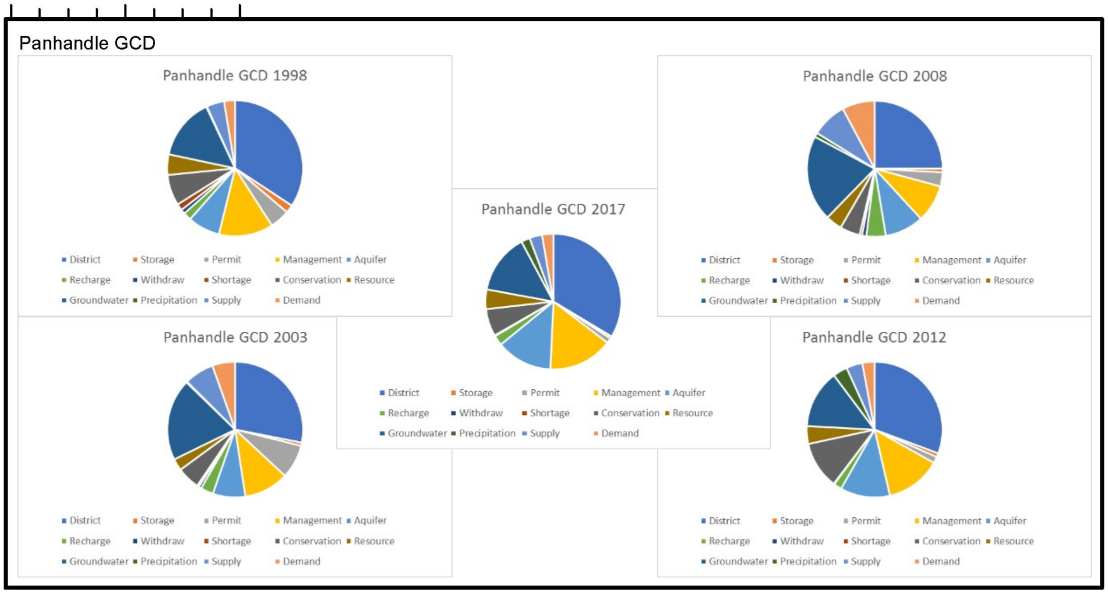
Northern Trinity GCD



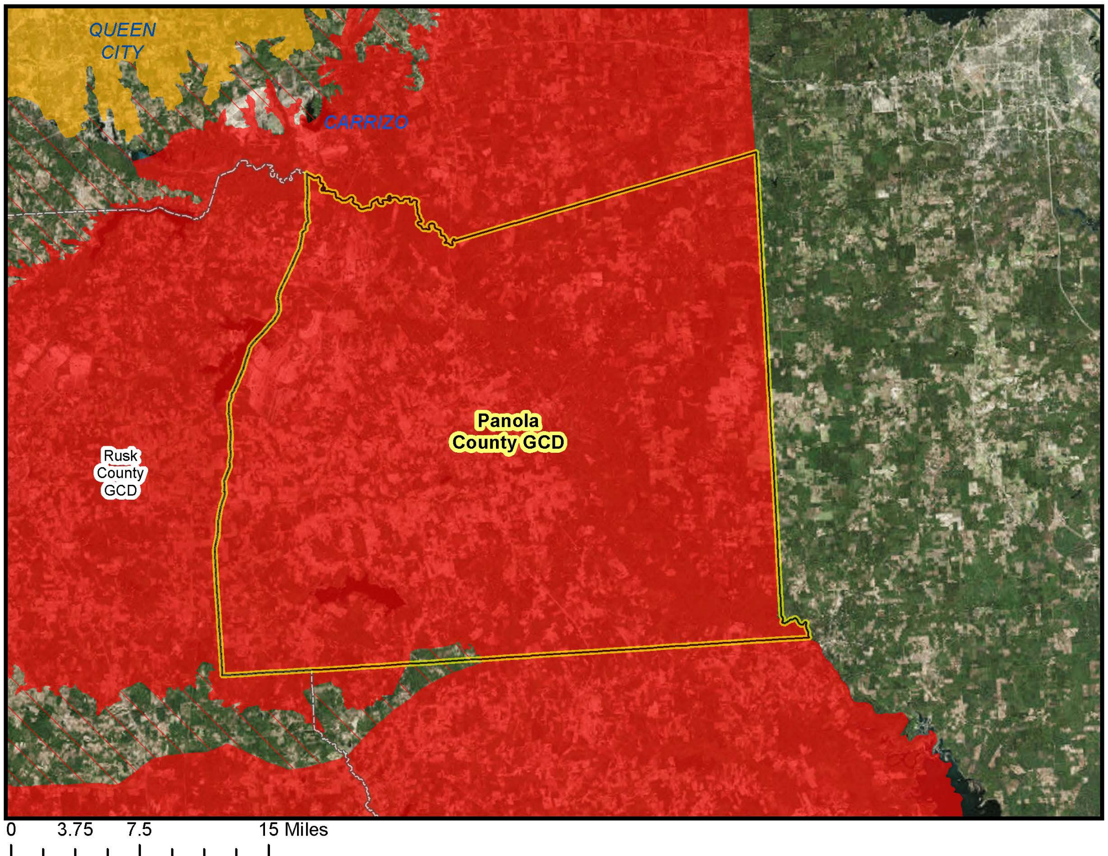


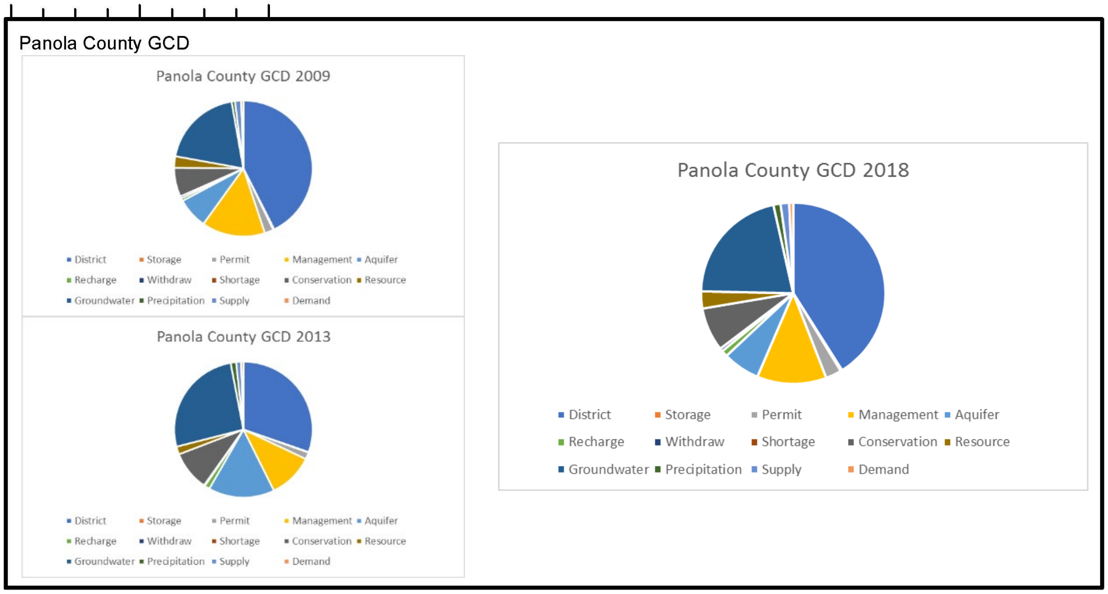
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



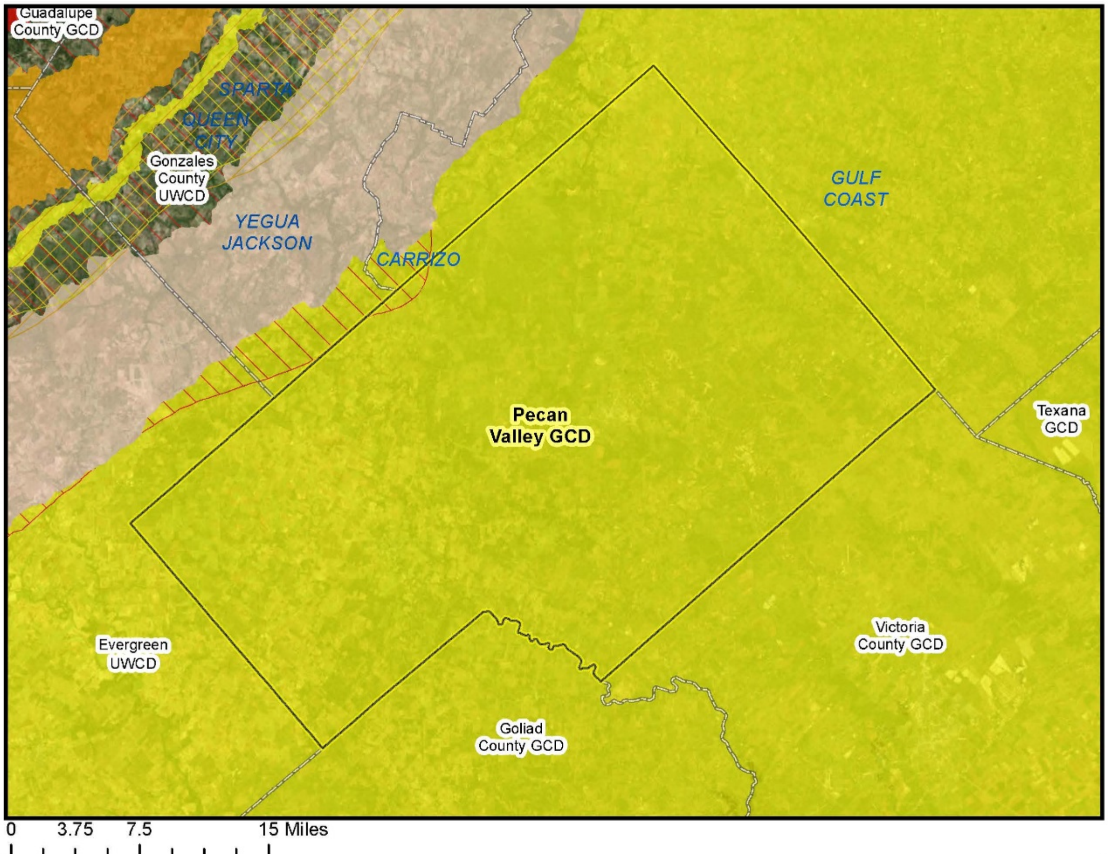


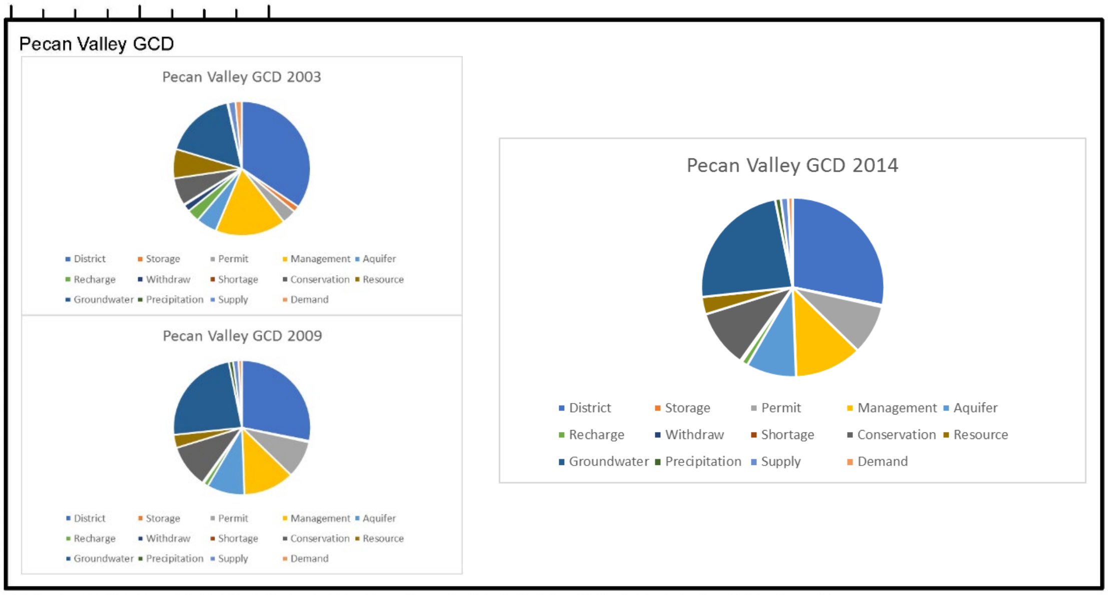
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



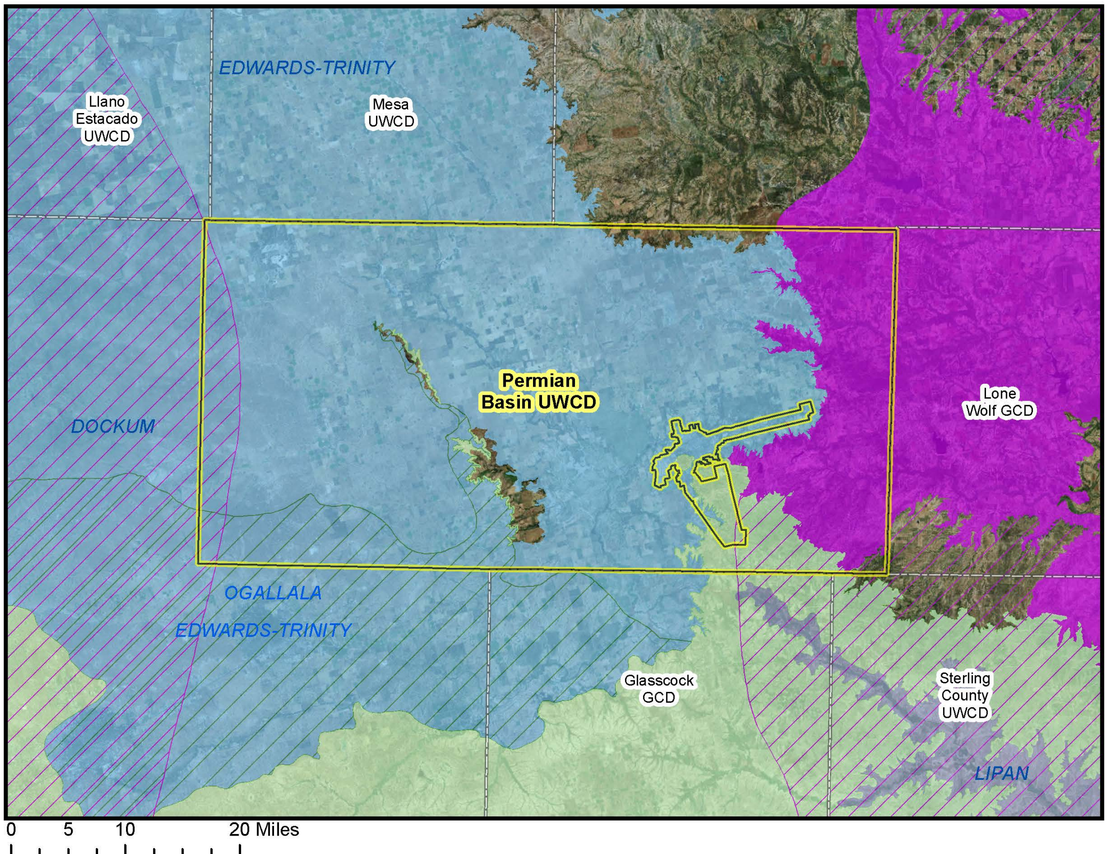


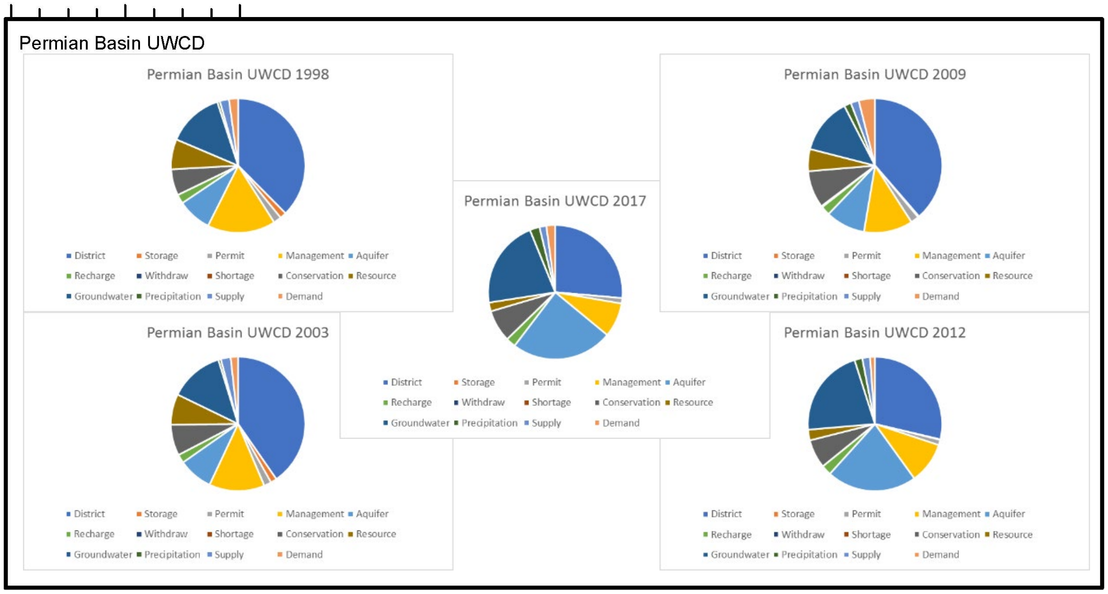
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



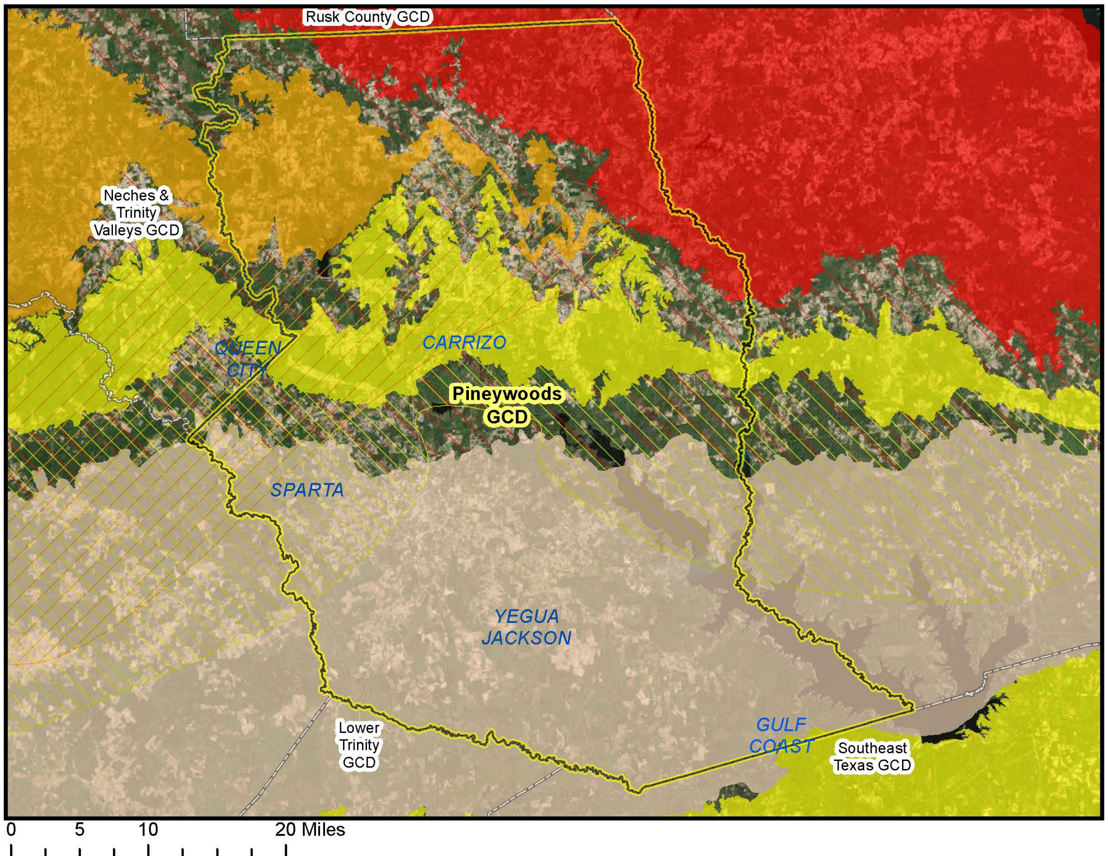


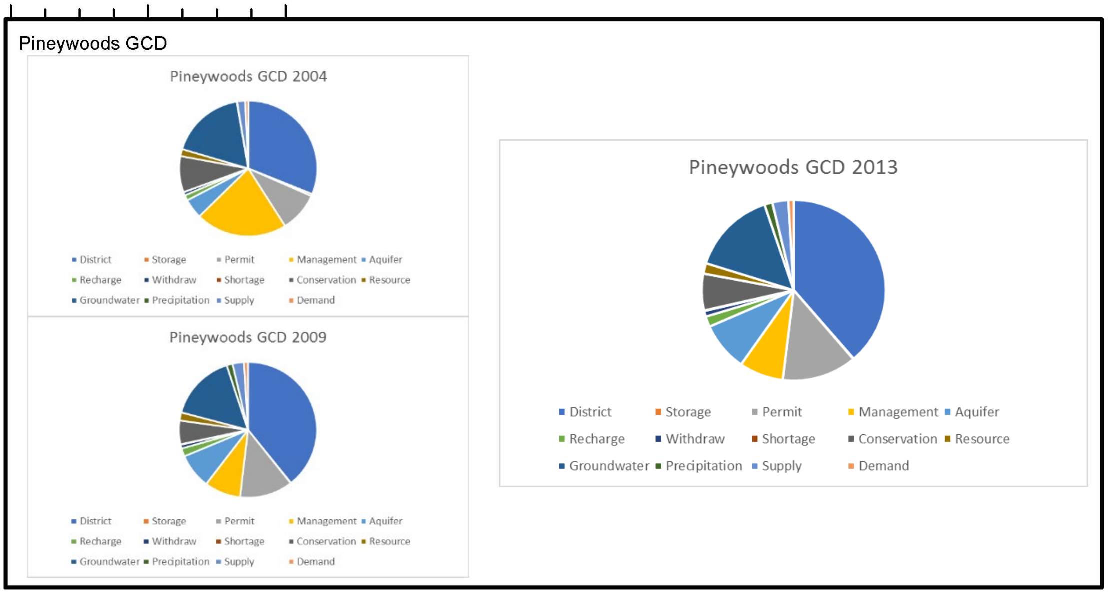
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



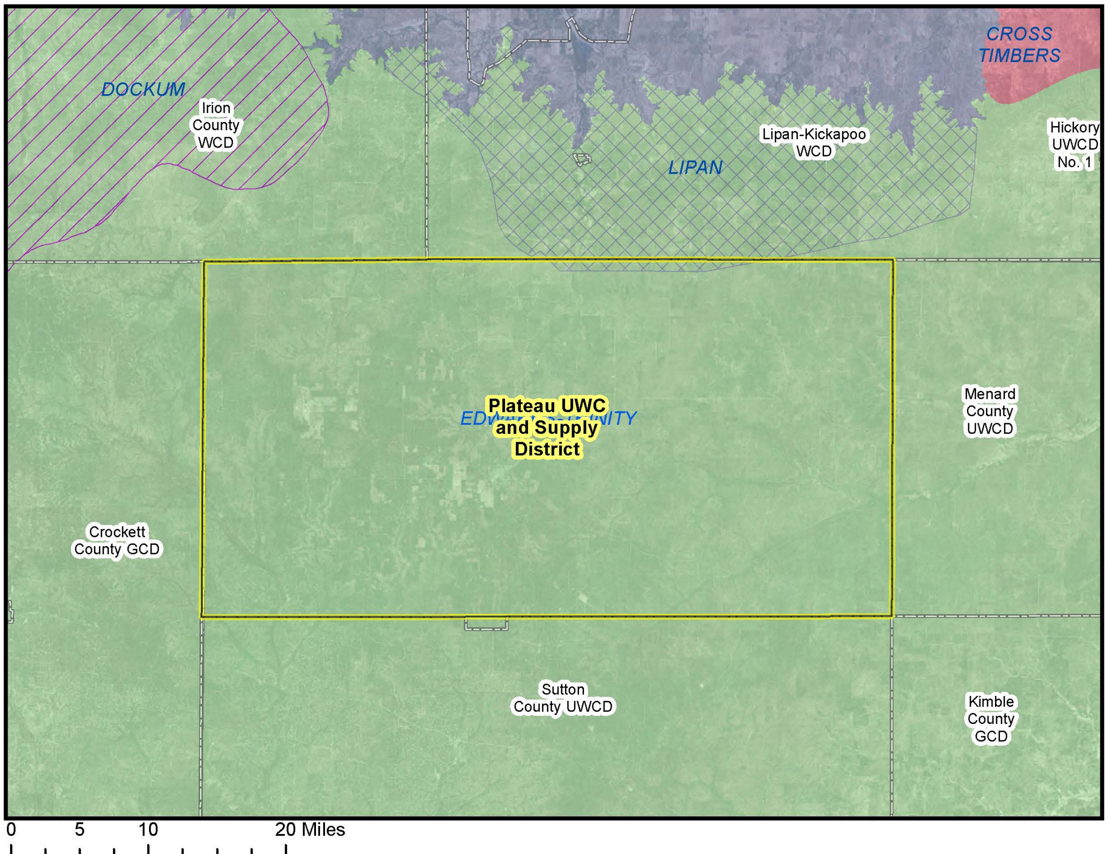


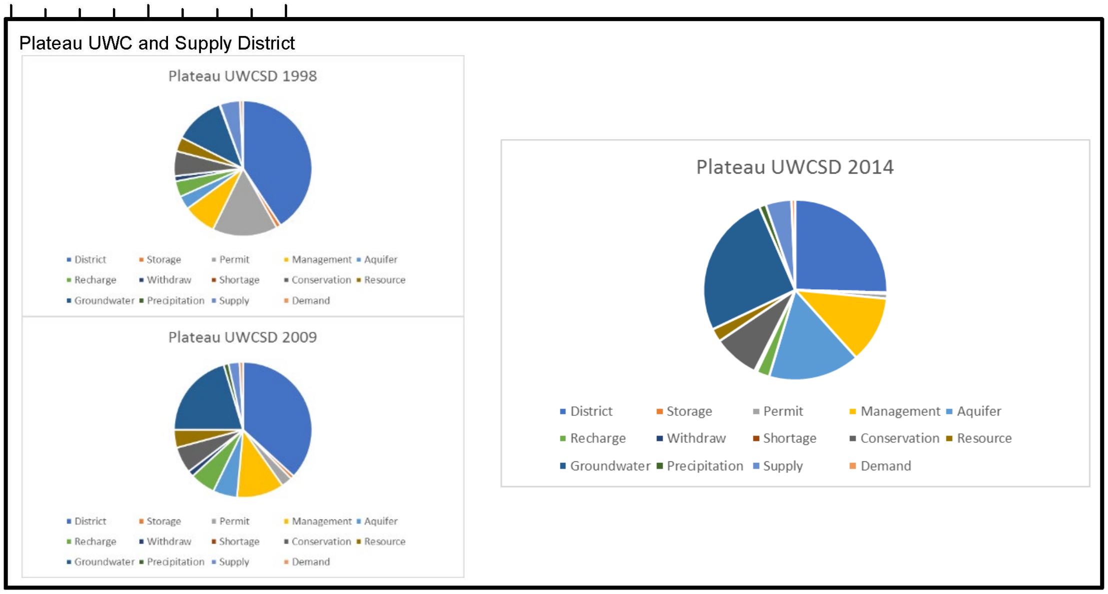
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



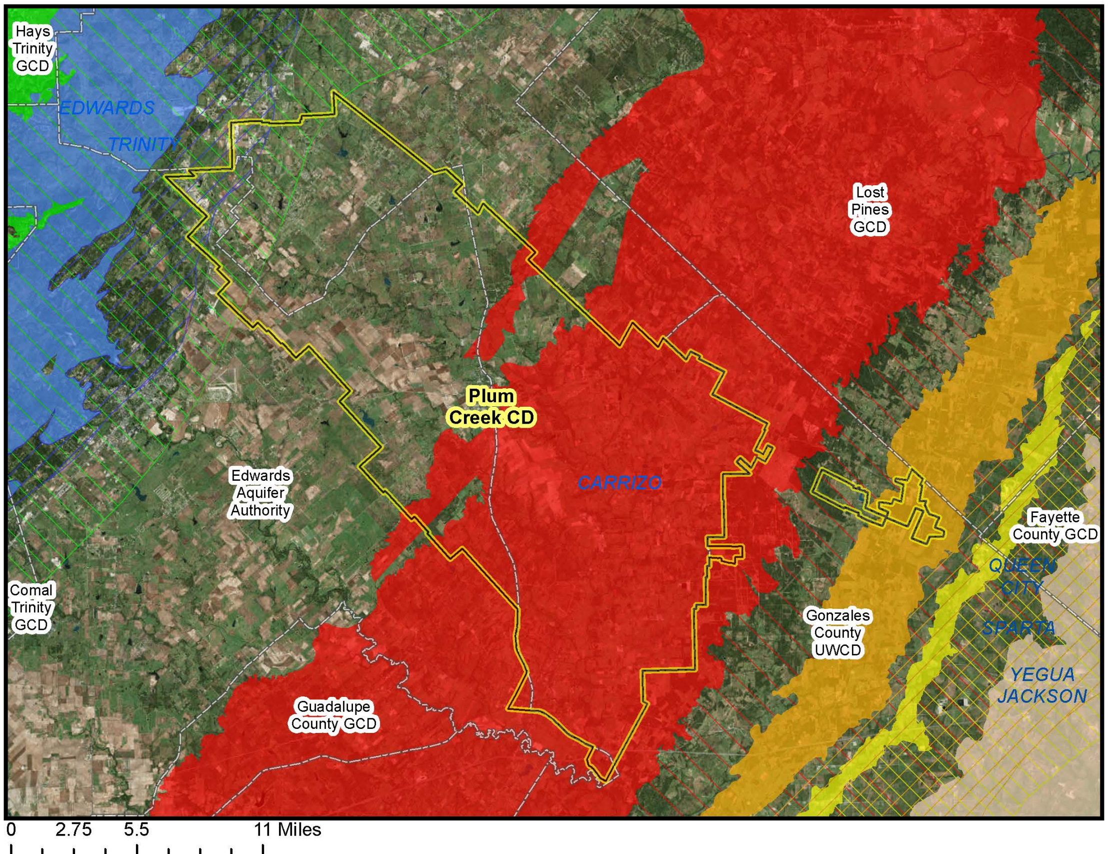


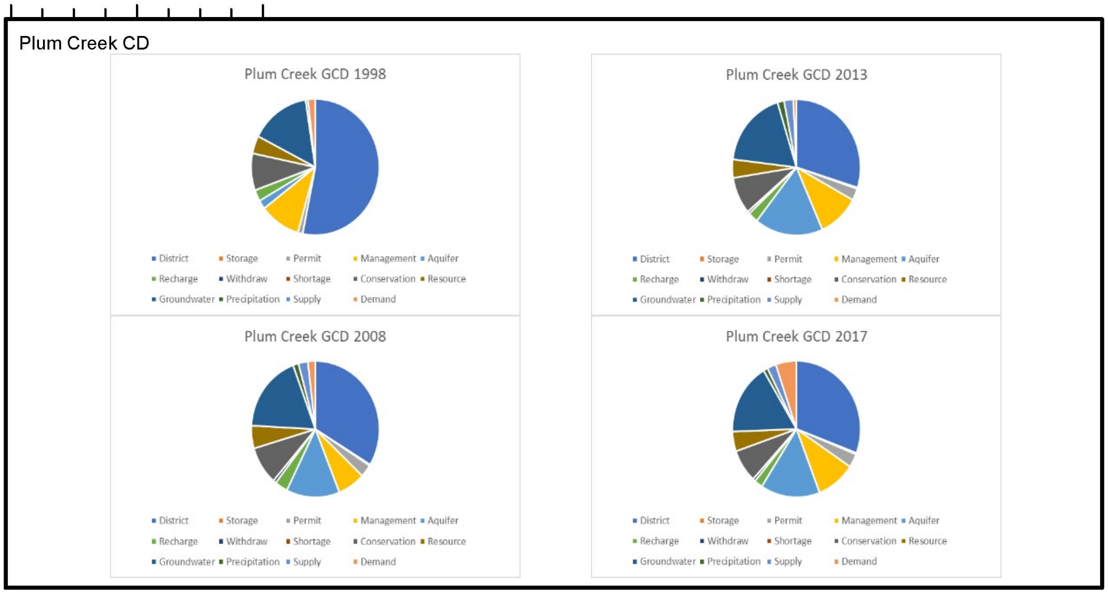
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



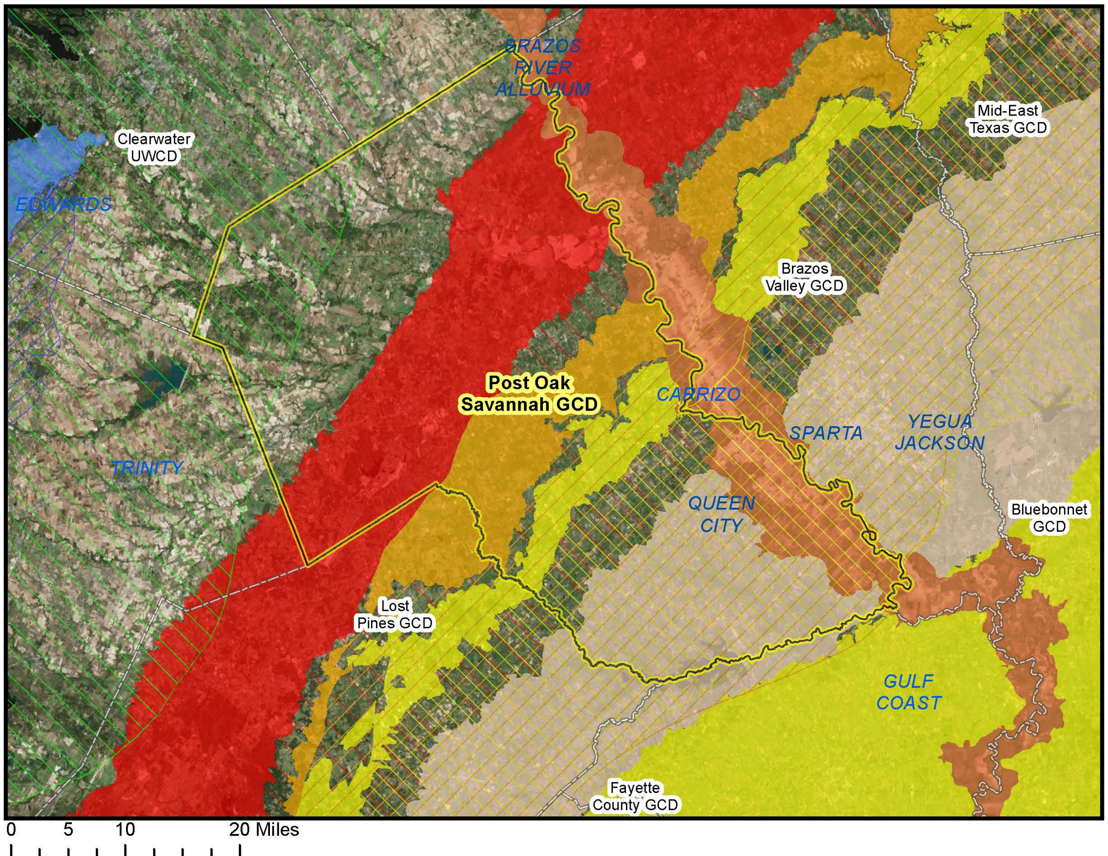


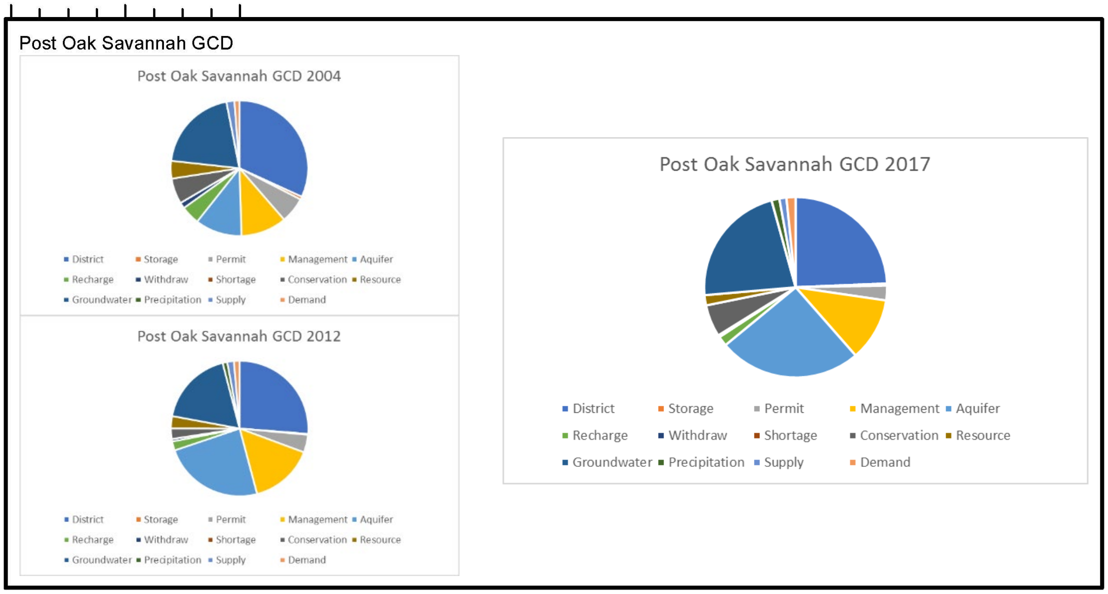
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



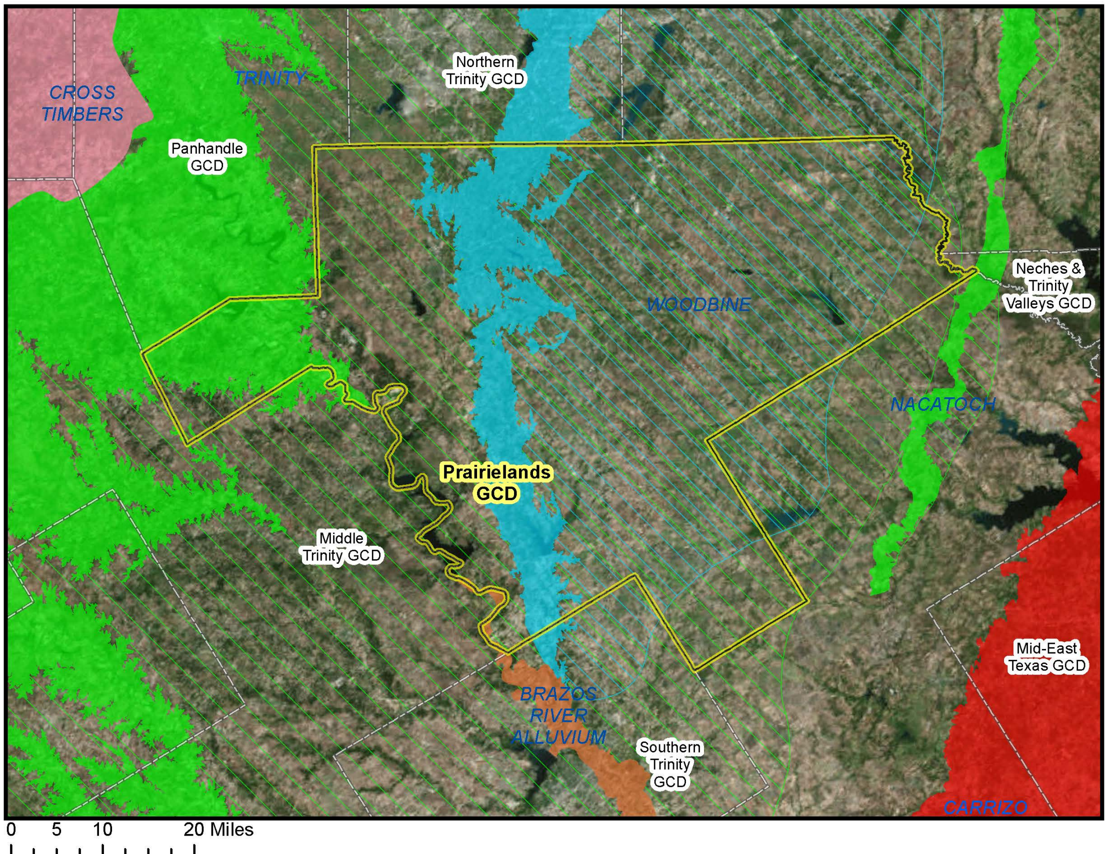


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

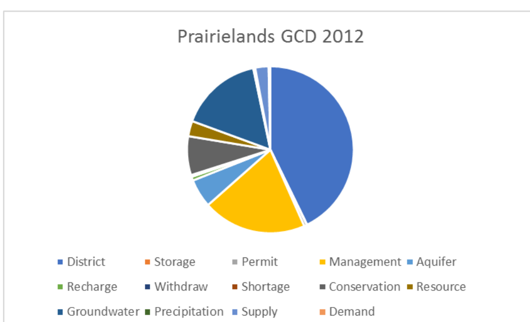


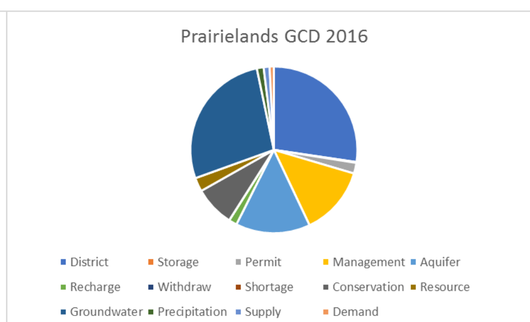


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

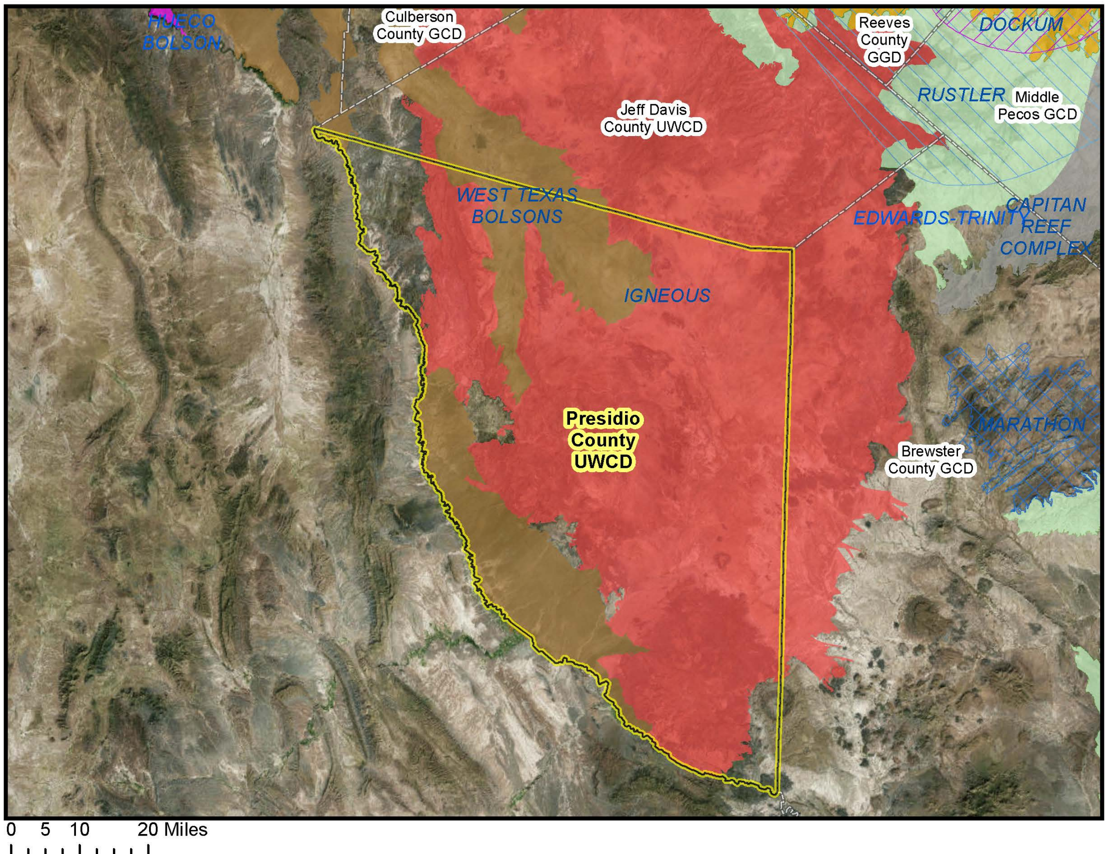


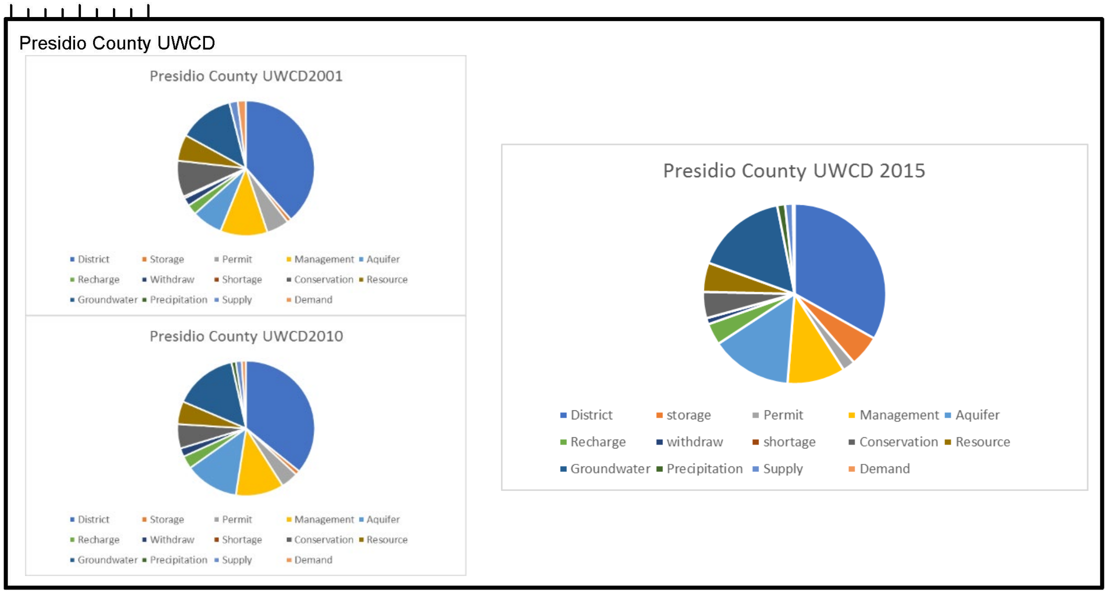
Prairielands GCD



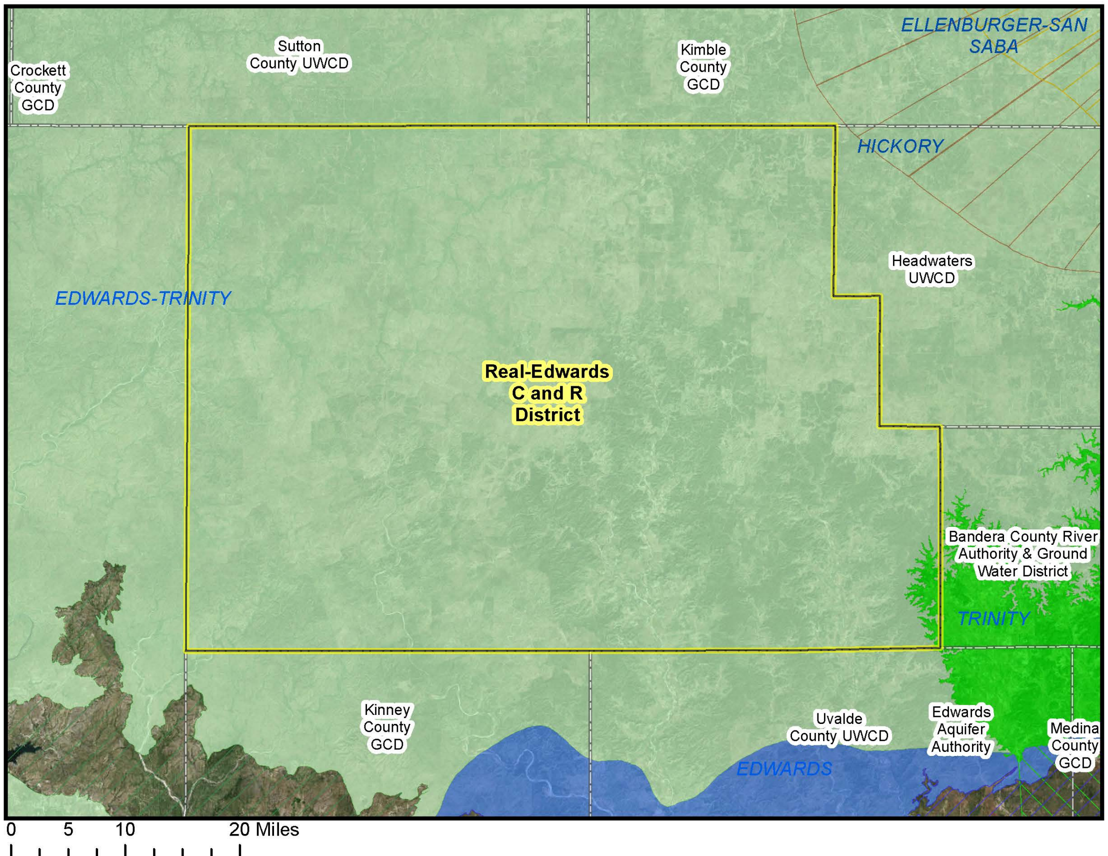


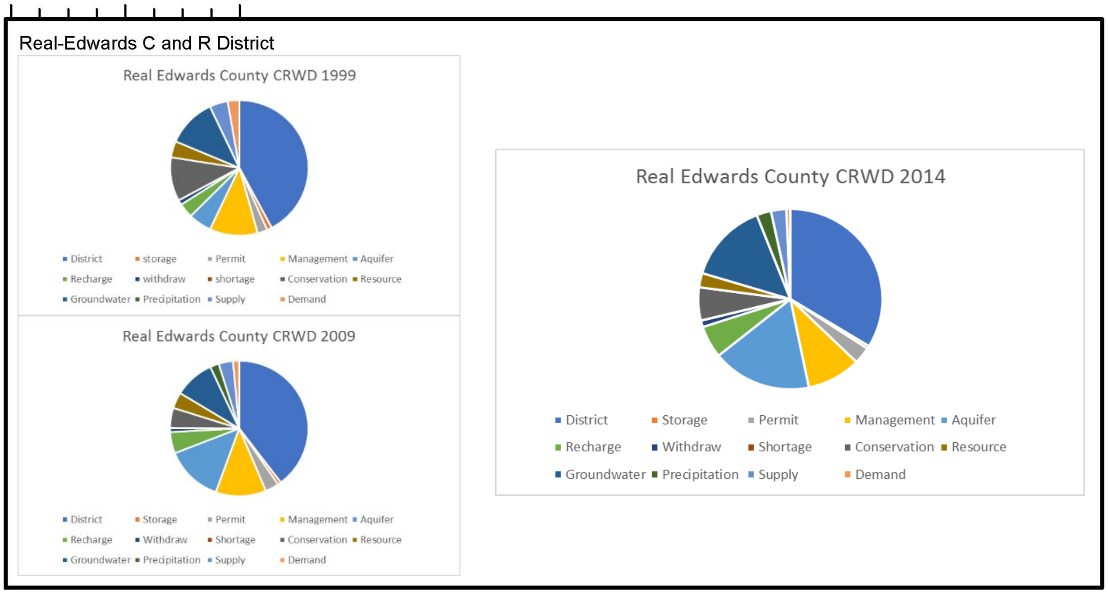
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



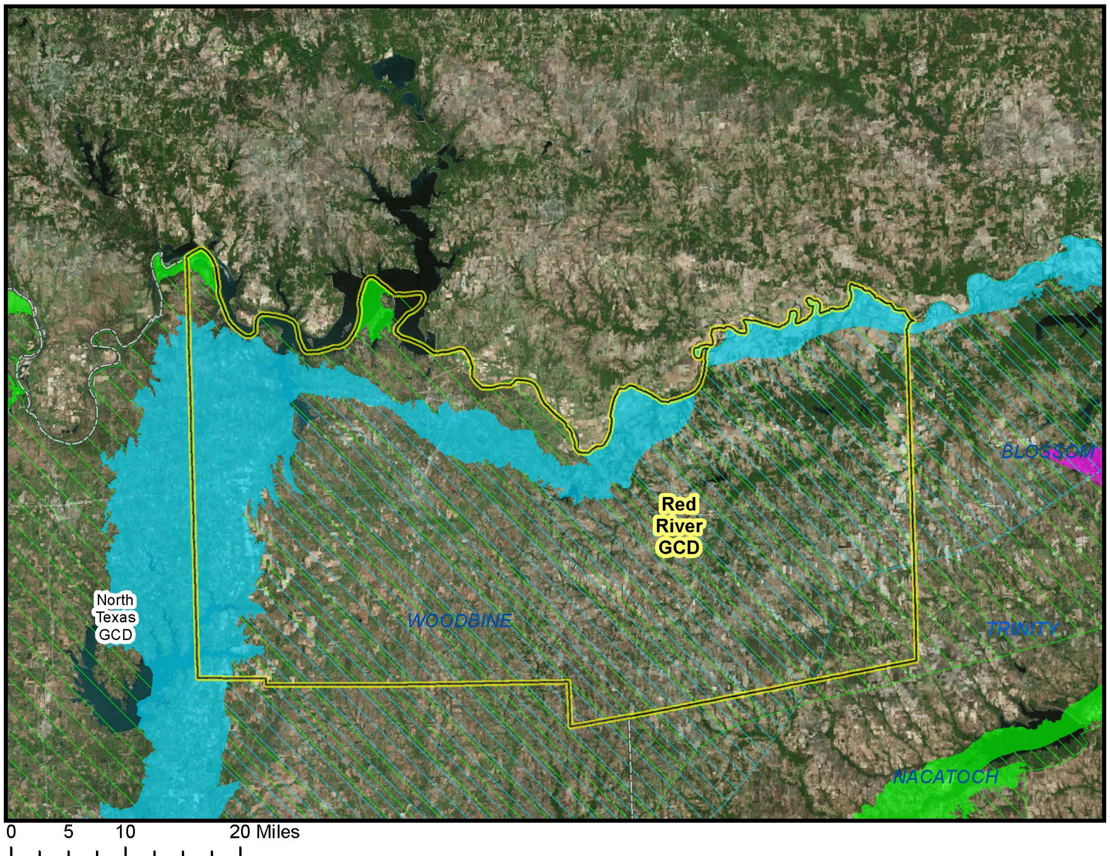


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

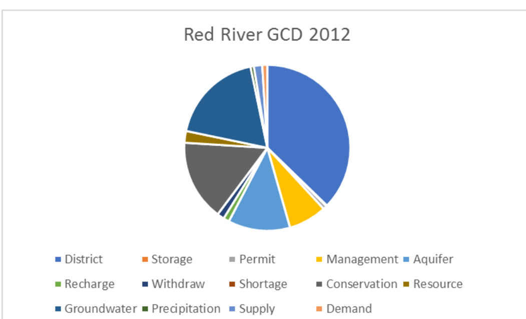


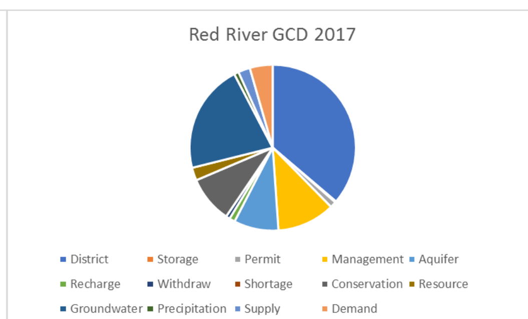


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

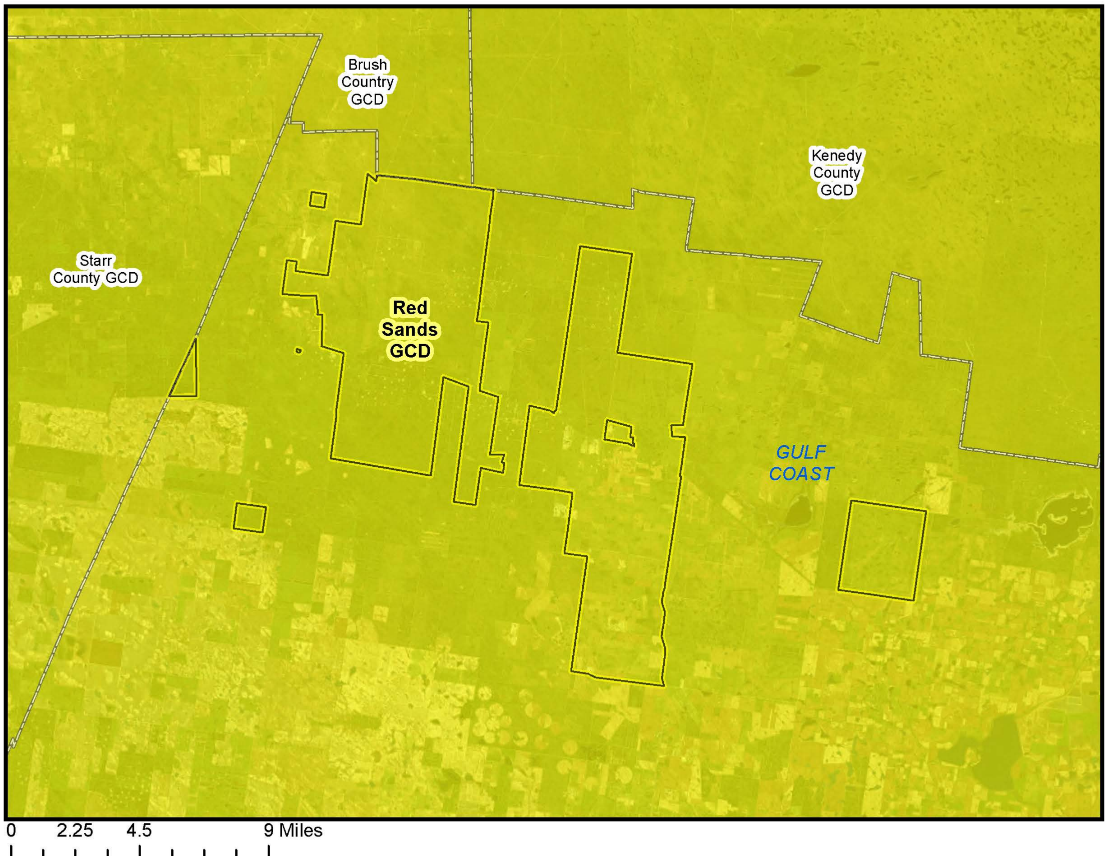


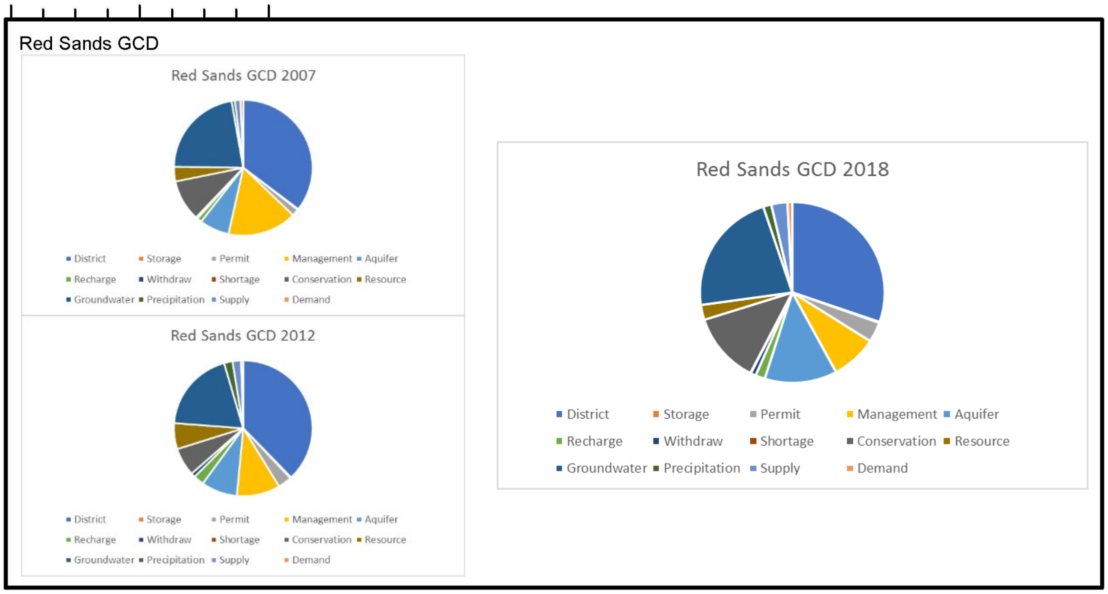
Red River GCD



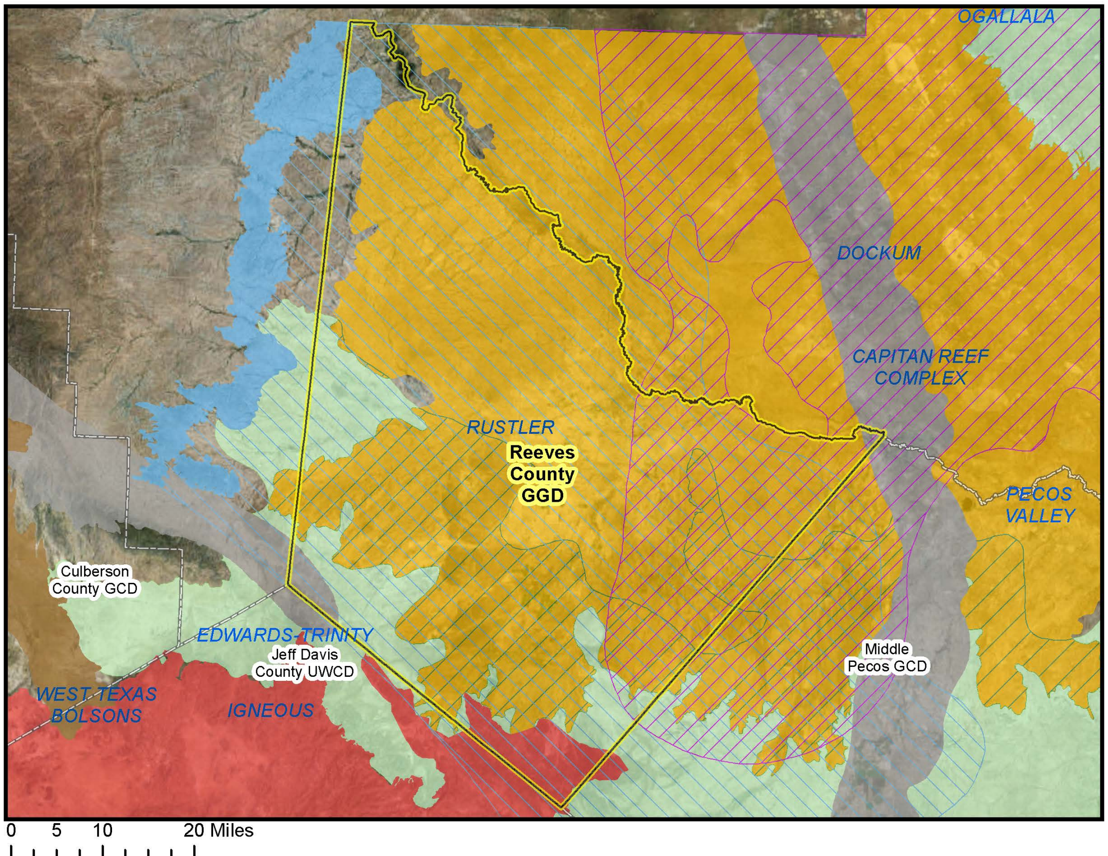


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

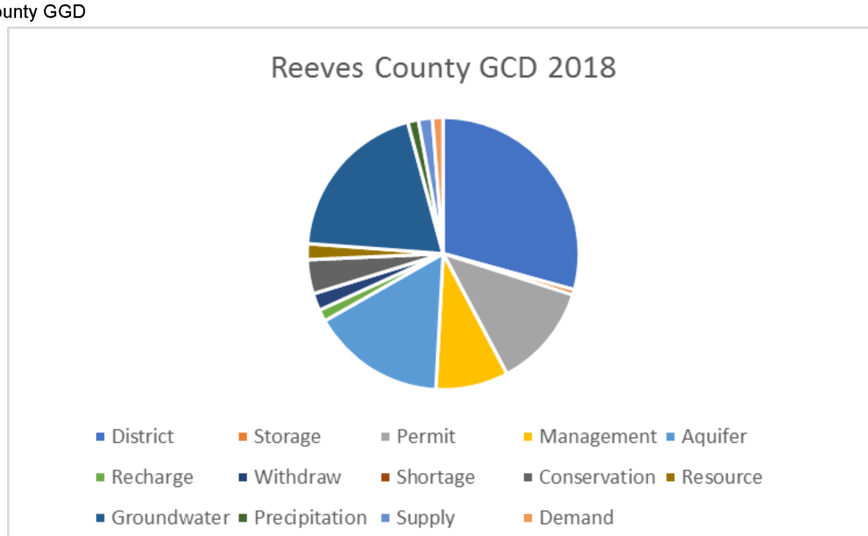




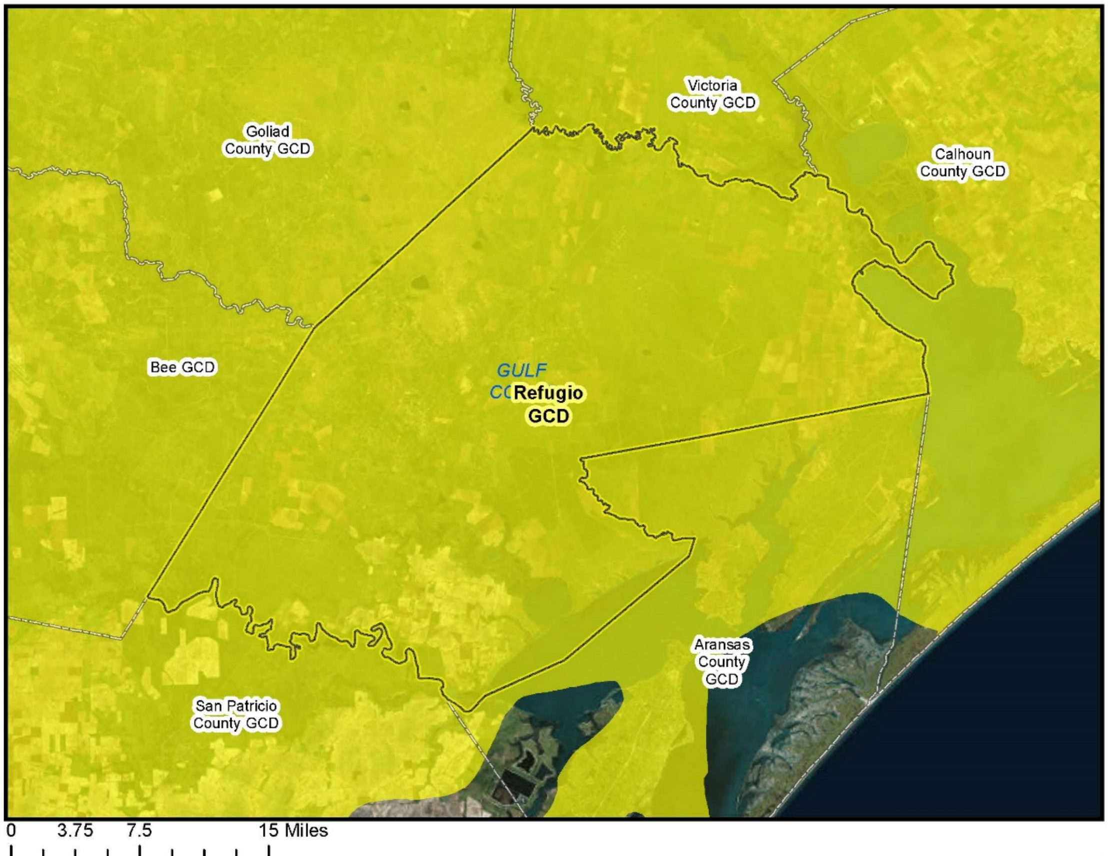
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

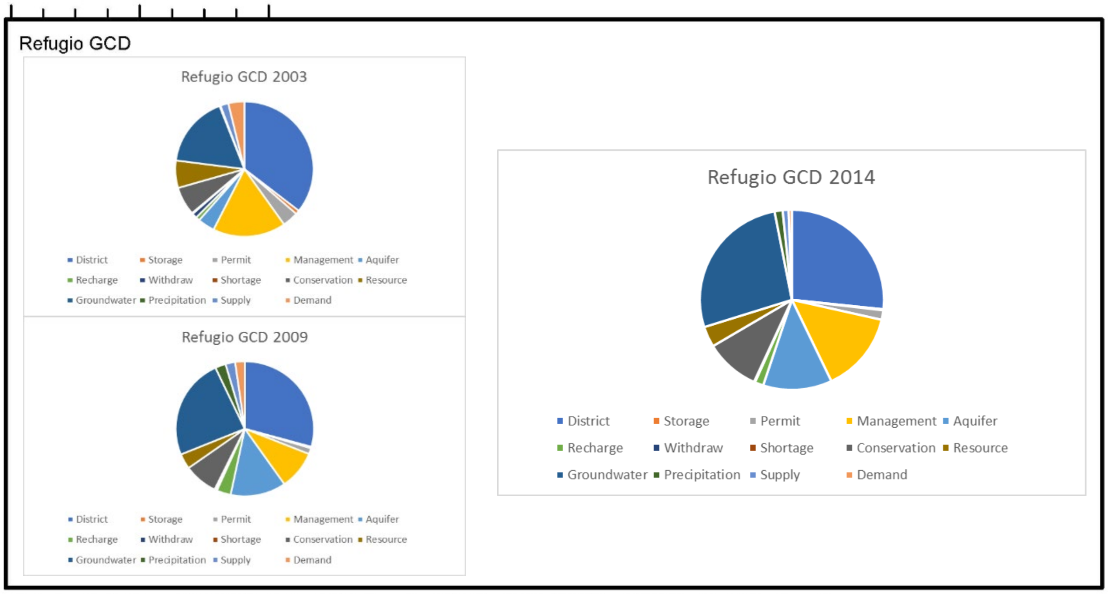


Reeves County GGD

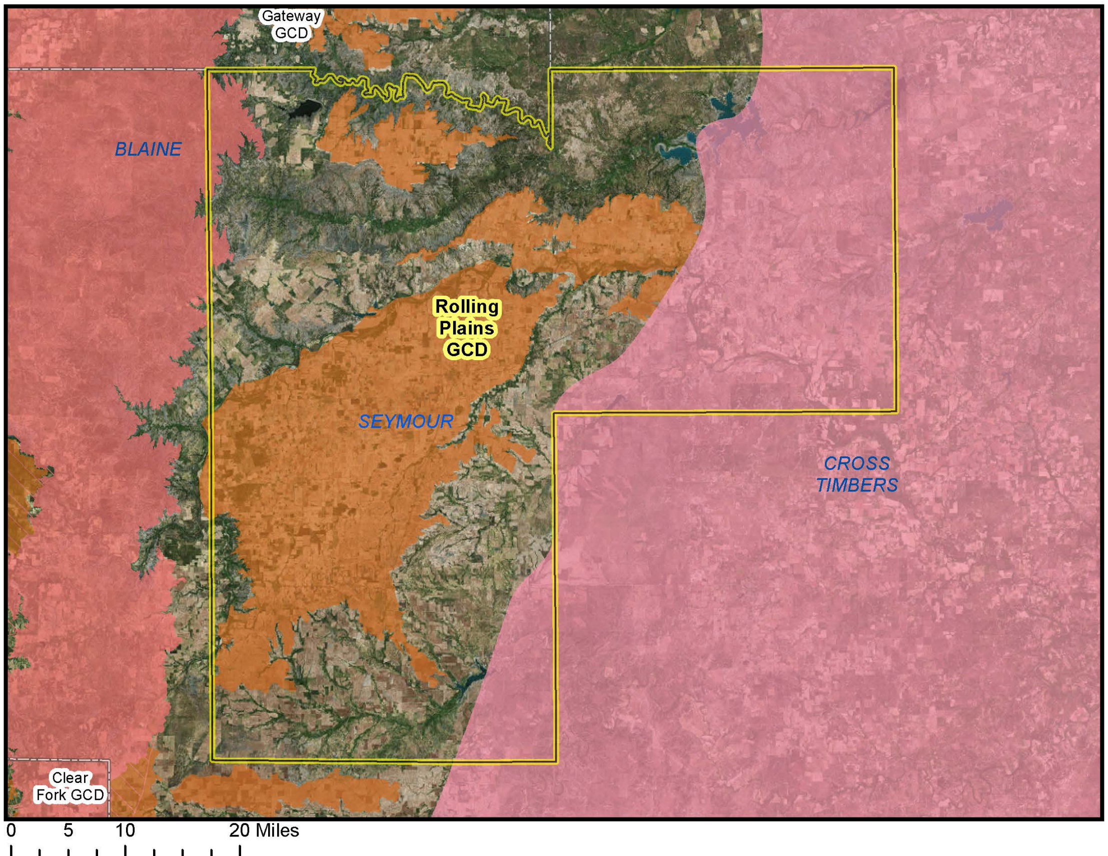


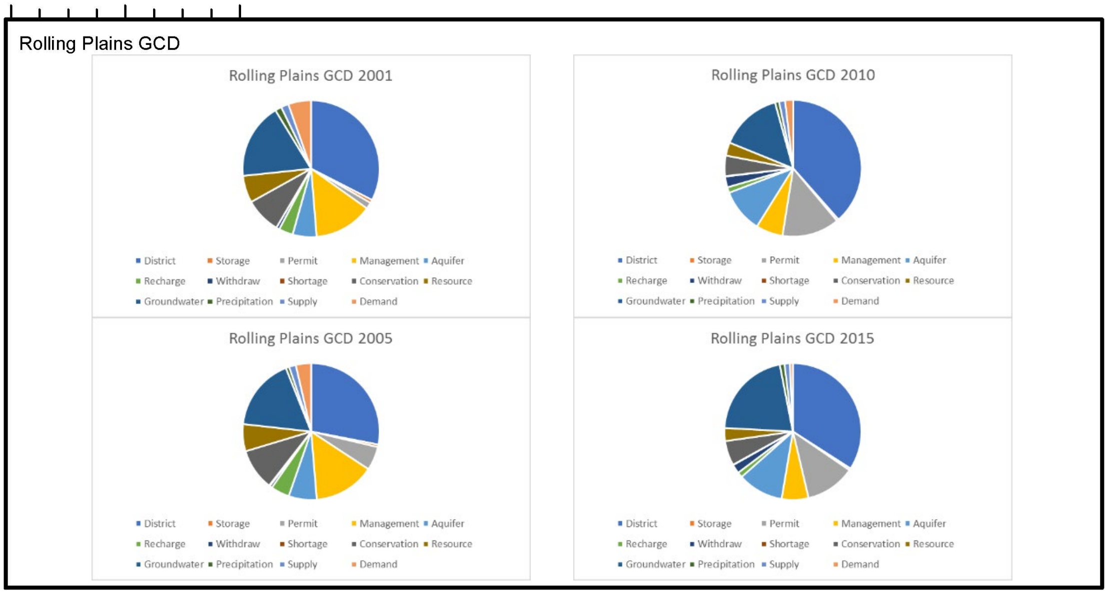
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



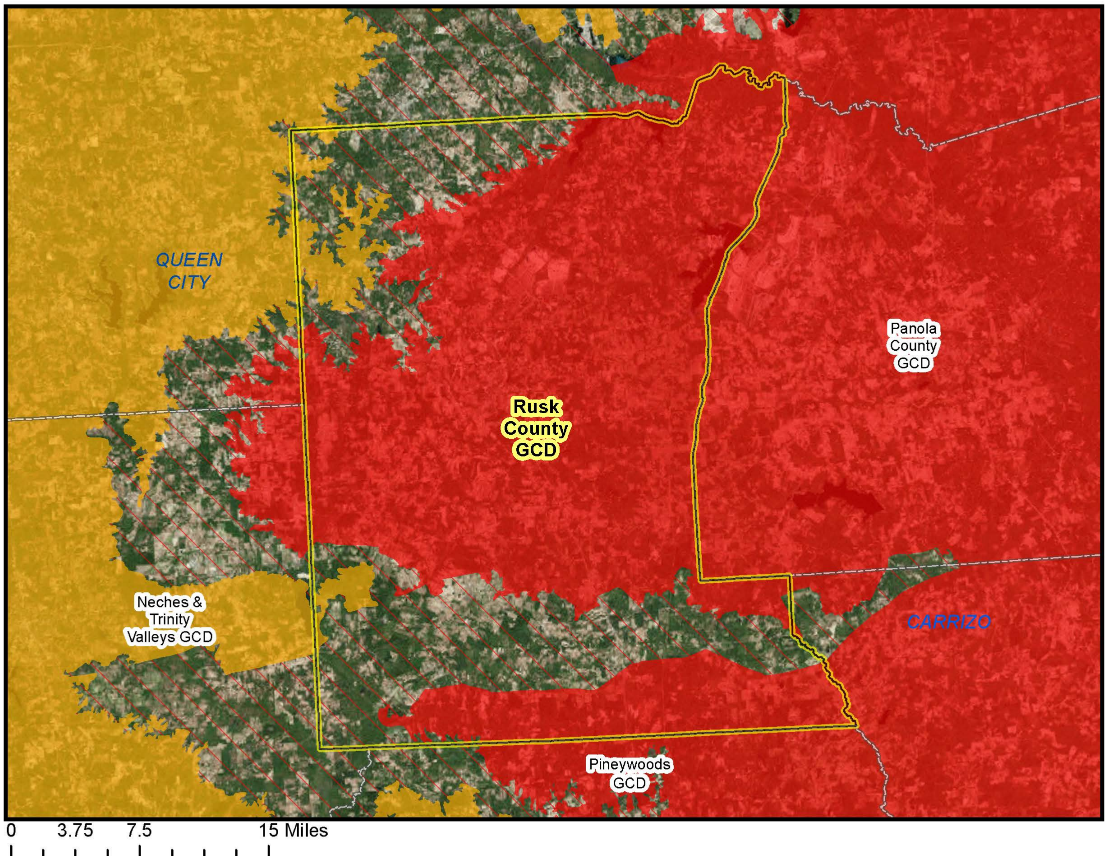


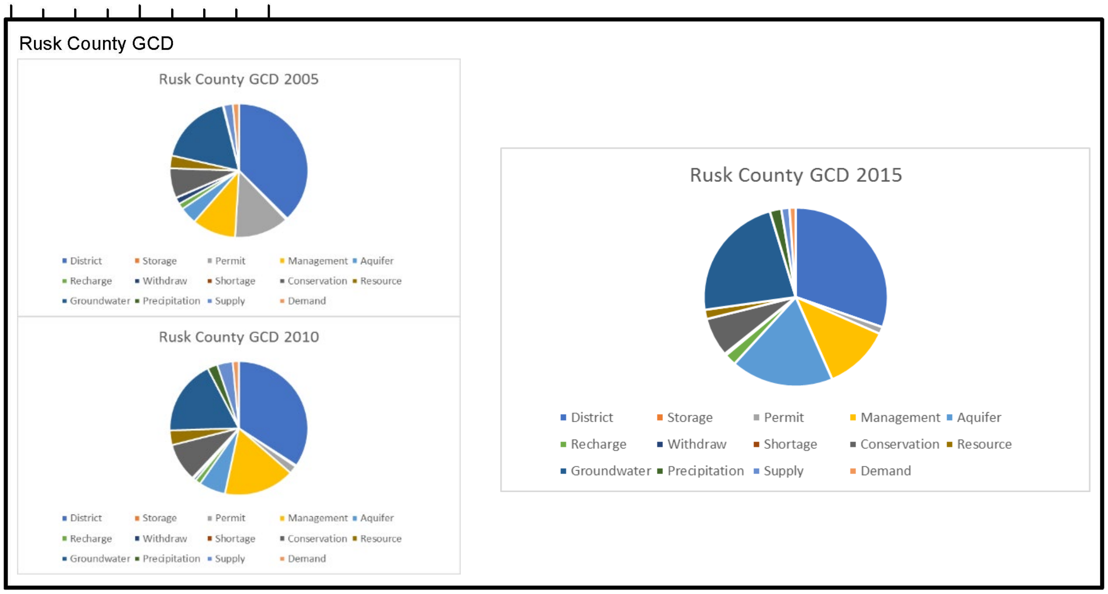
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



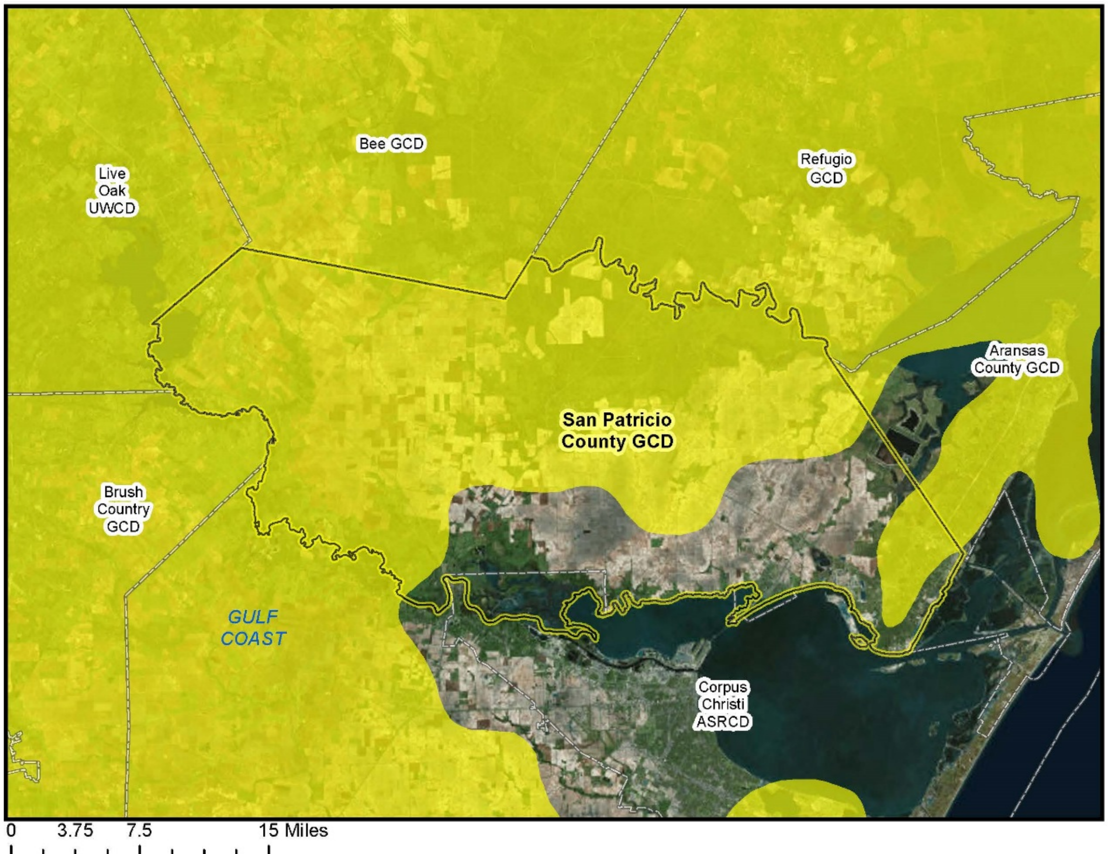


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

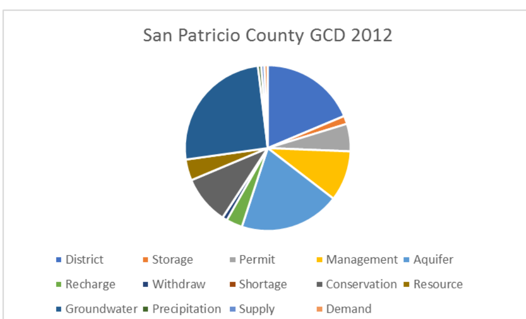


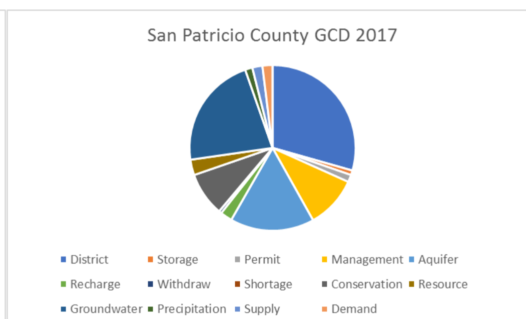


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

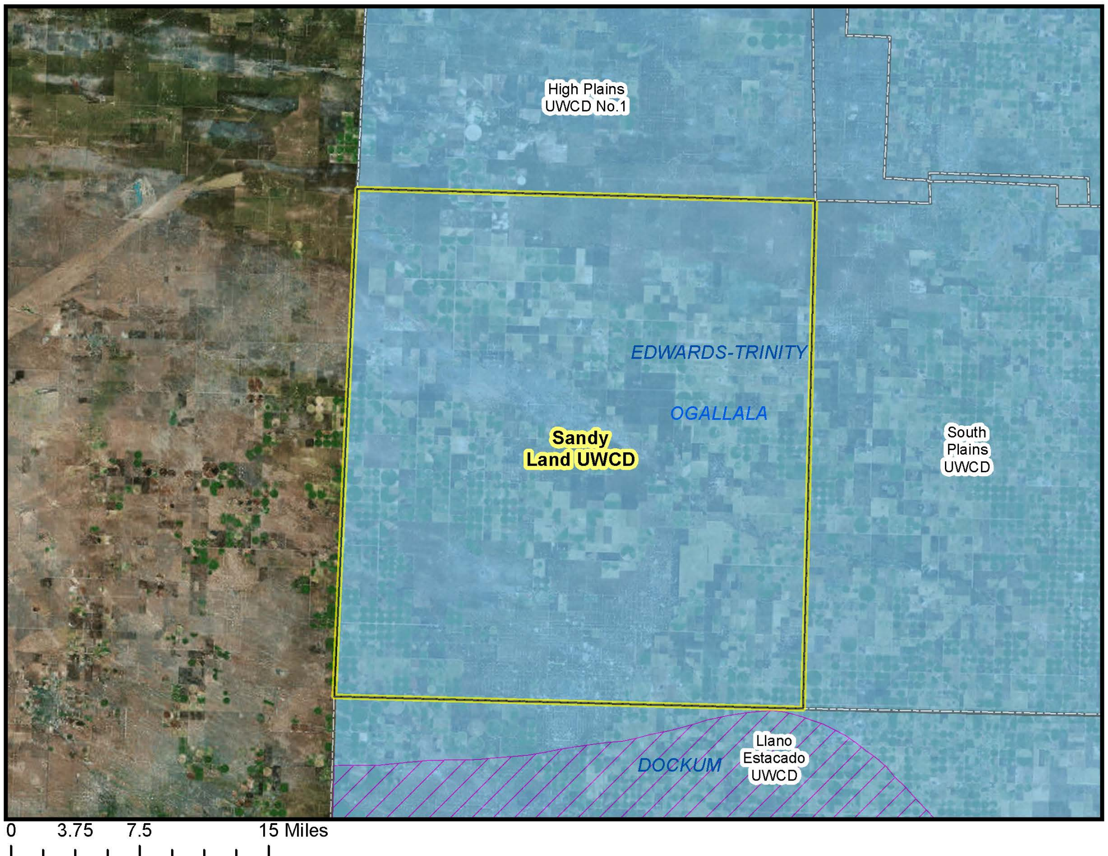


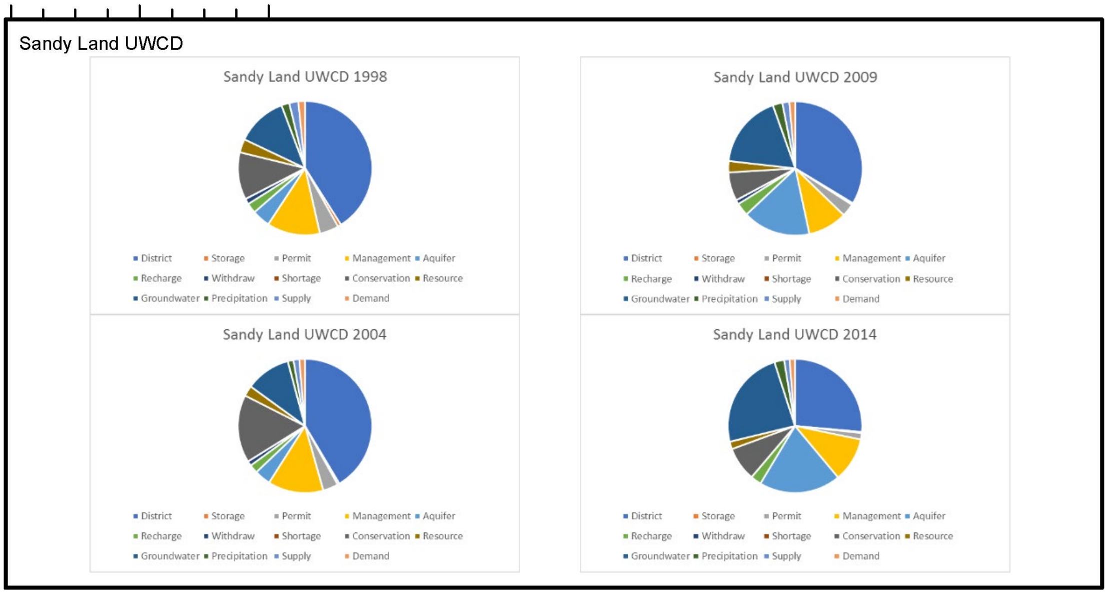
San Patricio County GCD



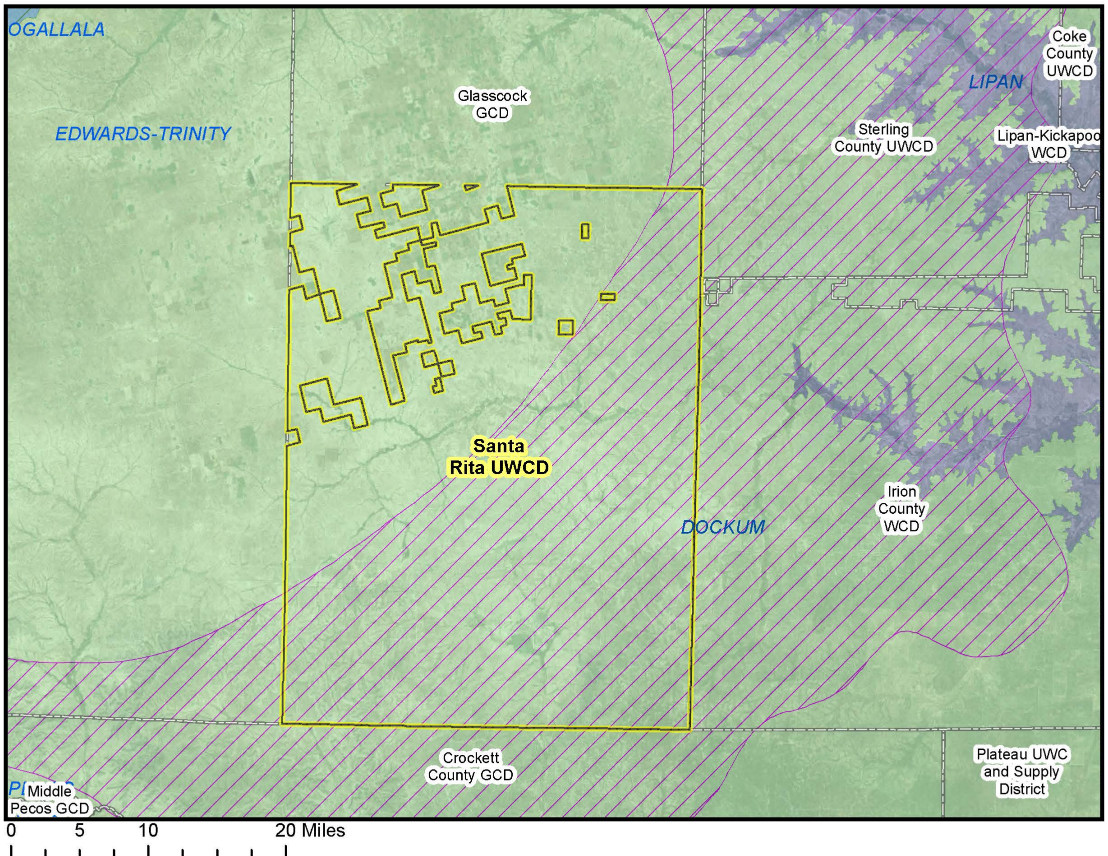


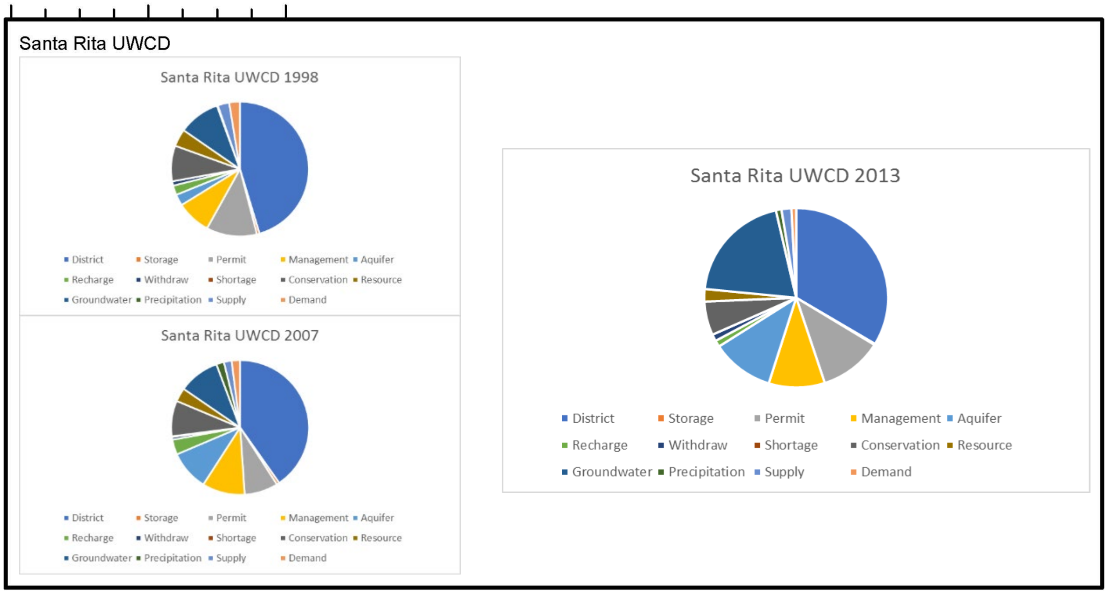
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



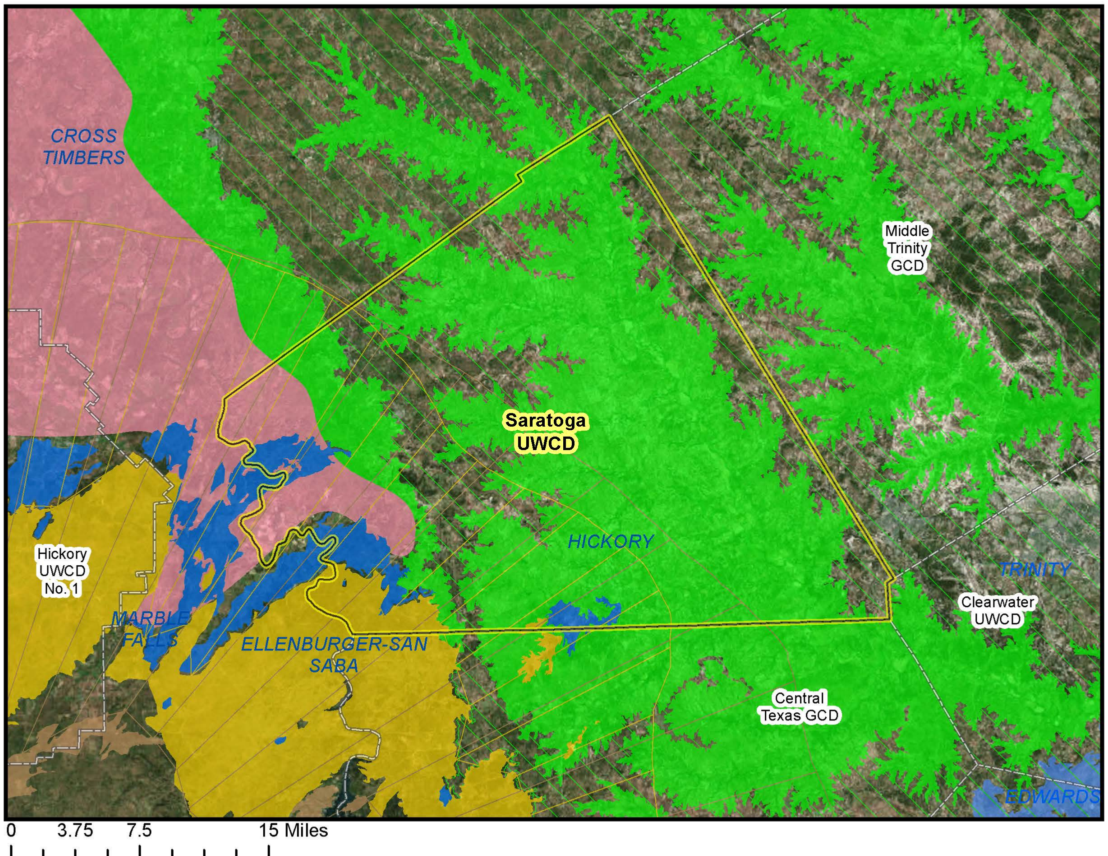


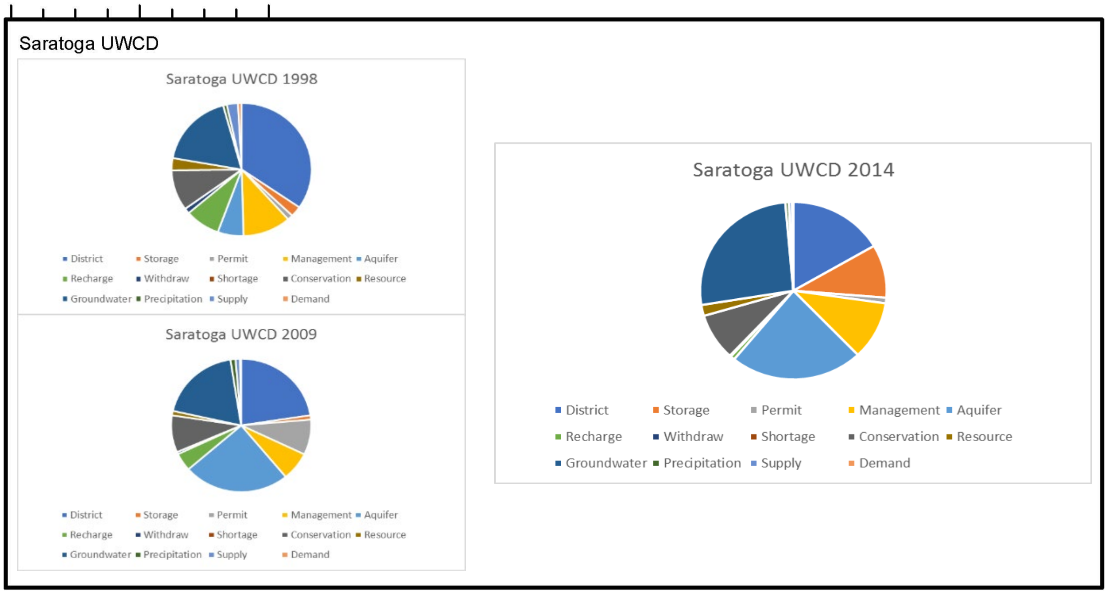
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



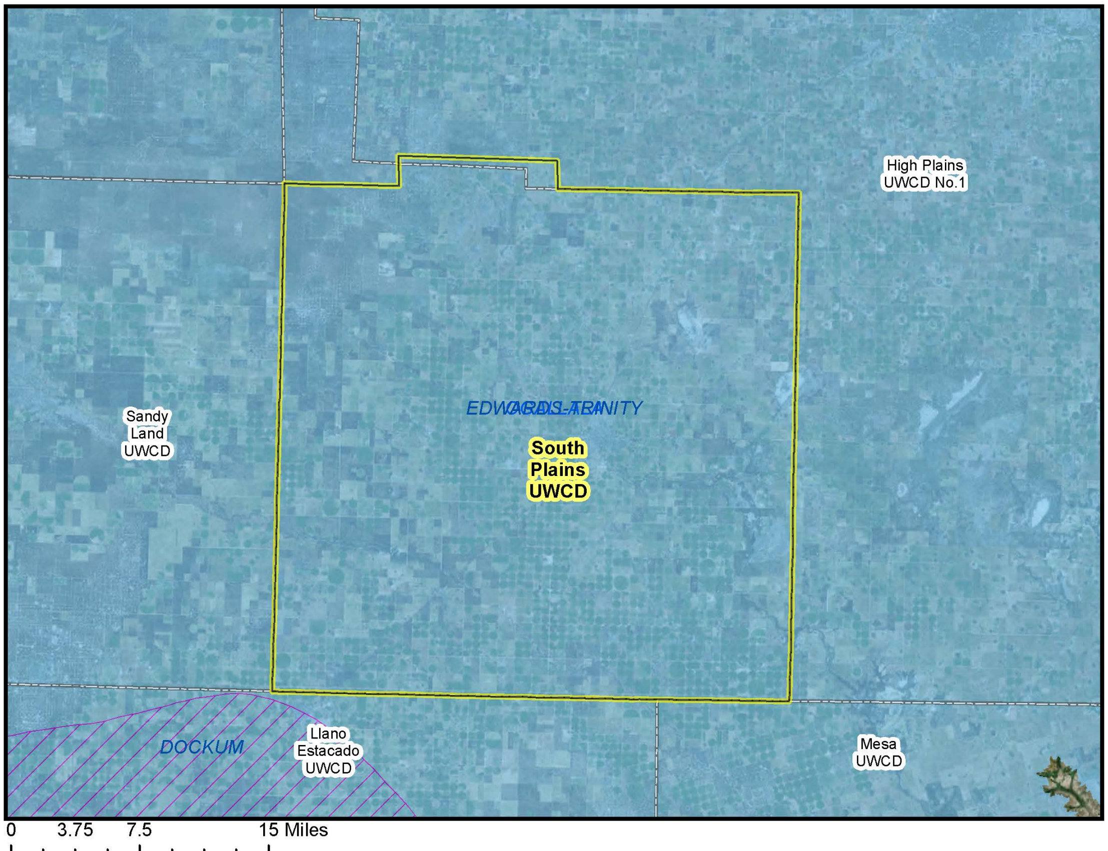


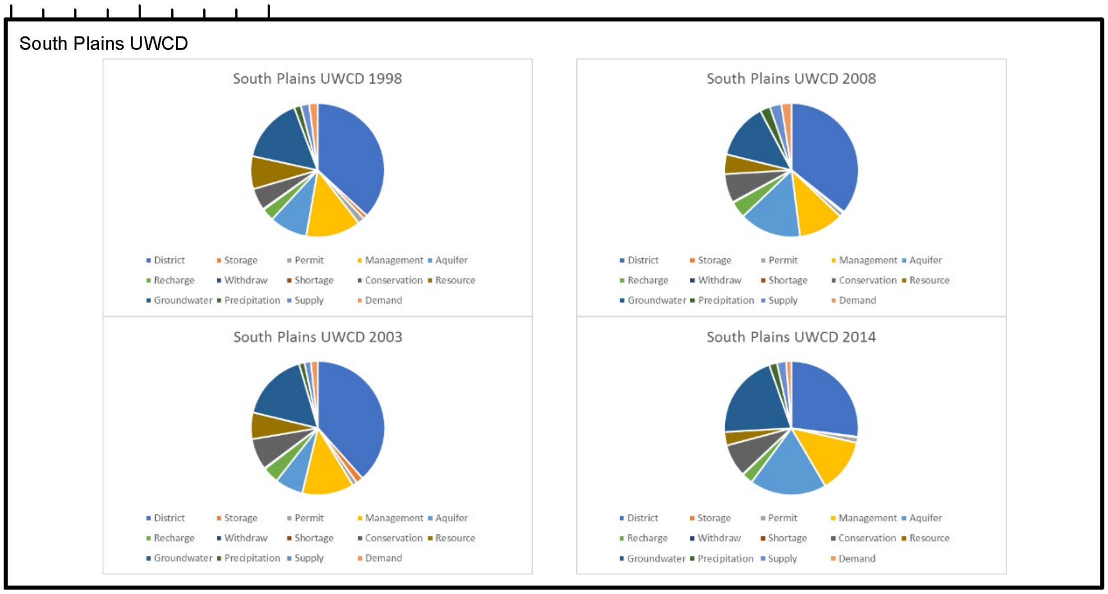
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



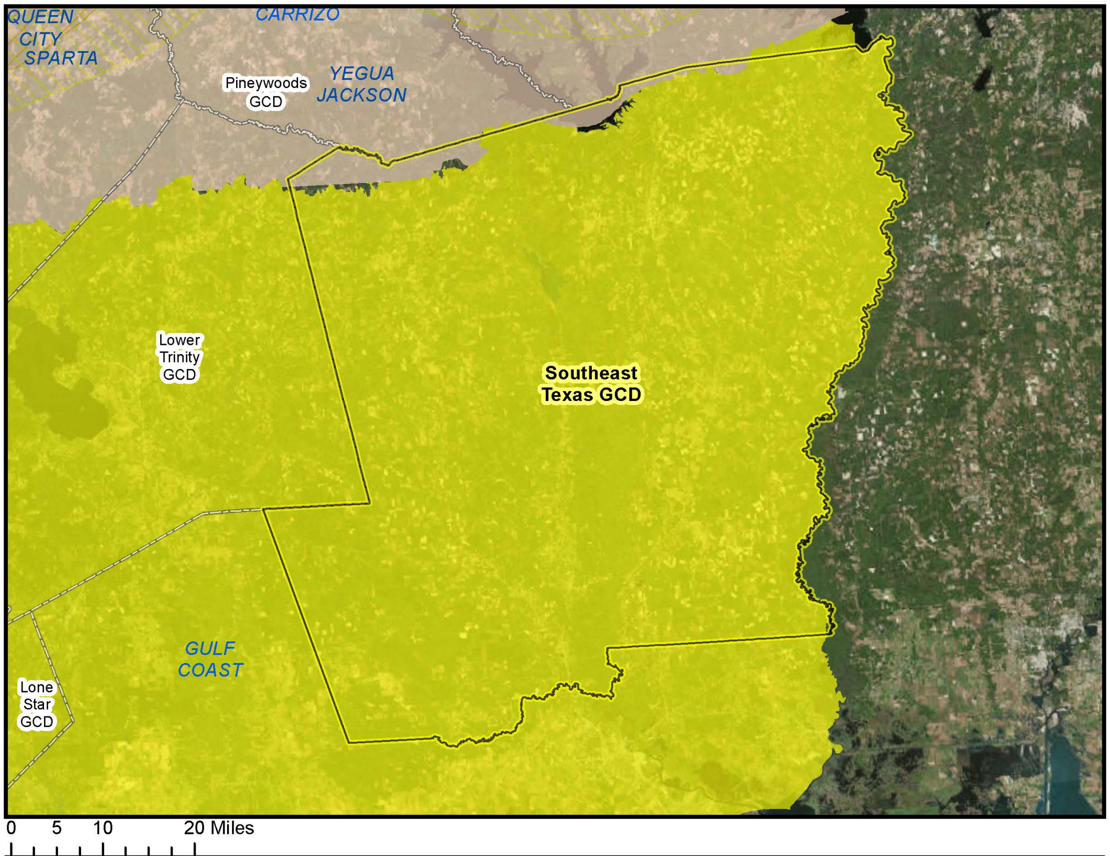


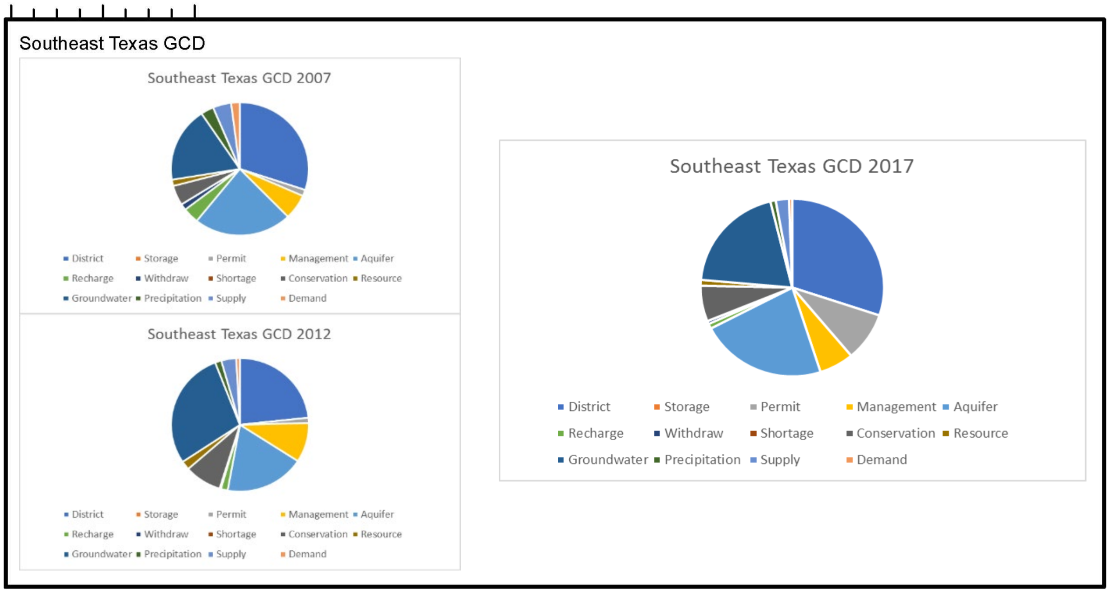
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



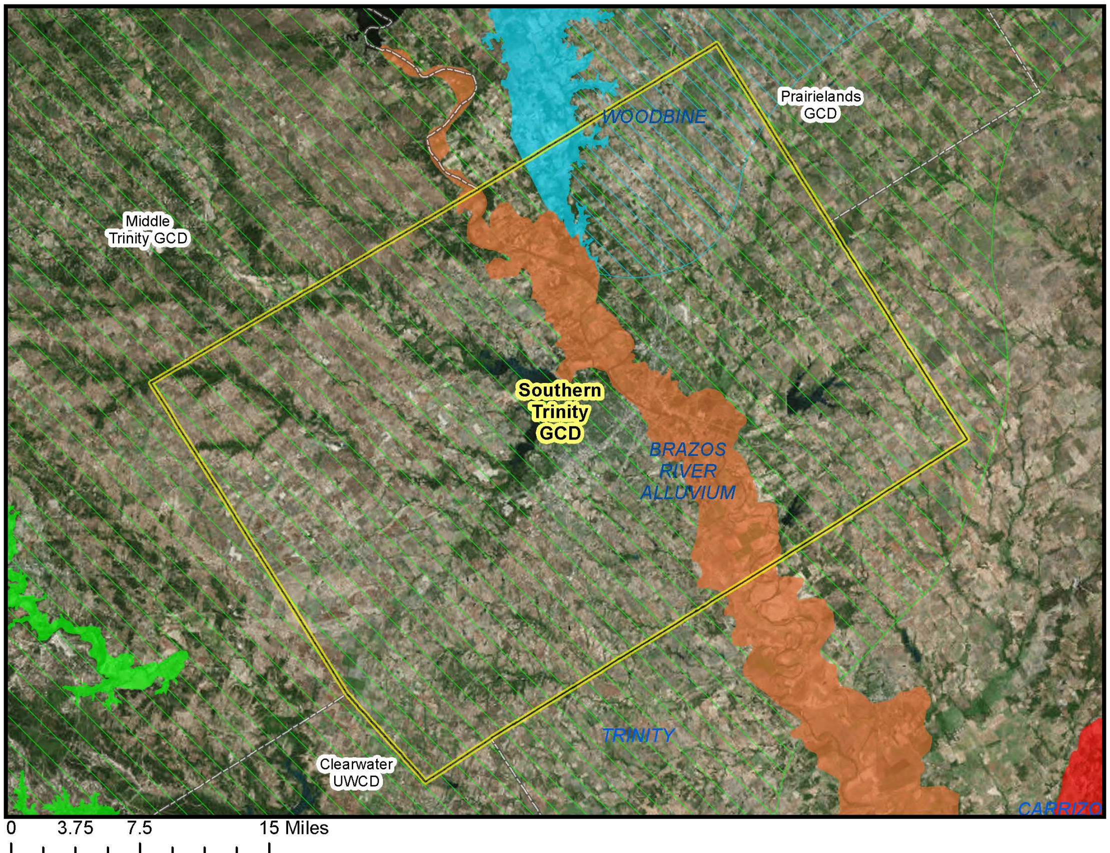


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

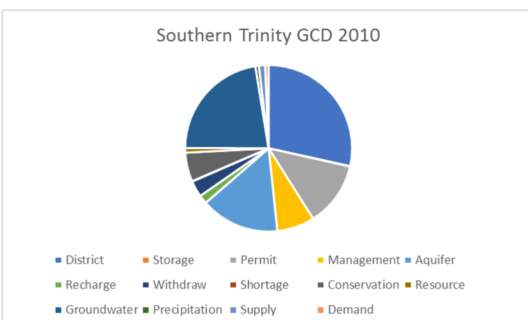


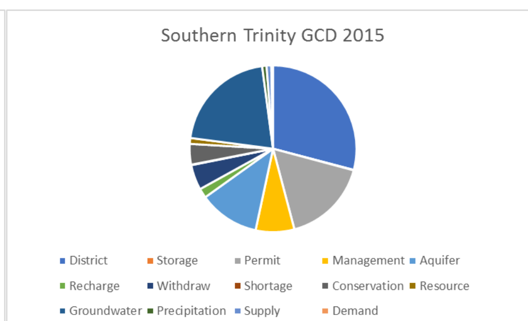


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

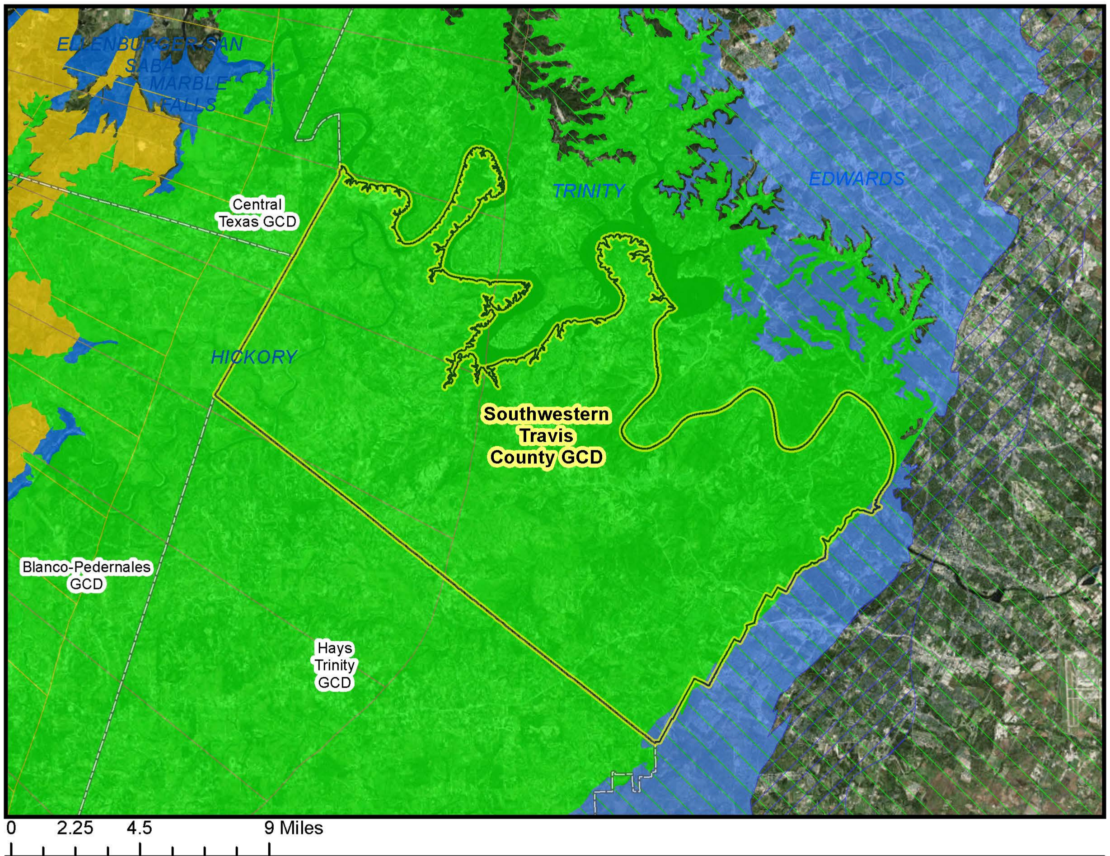


Southern Trinity GCD





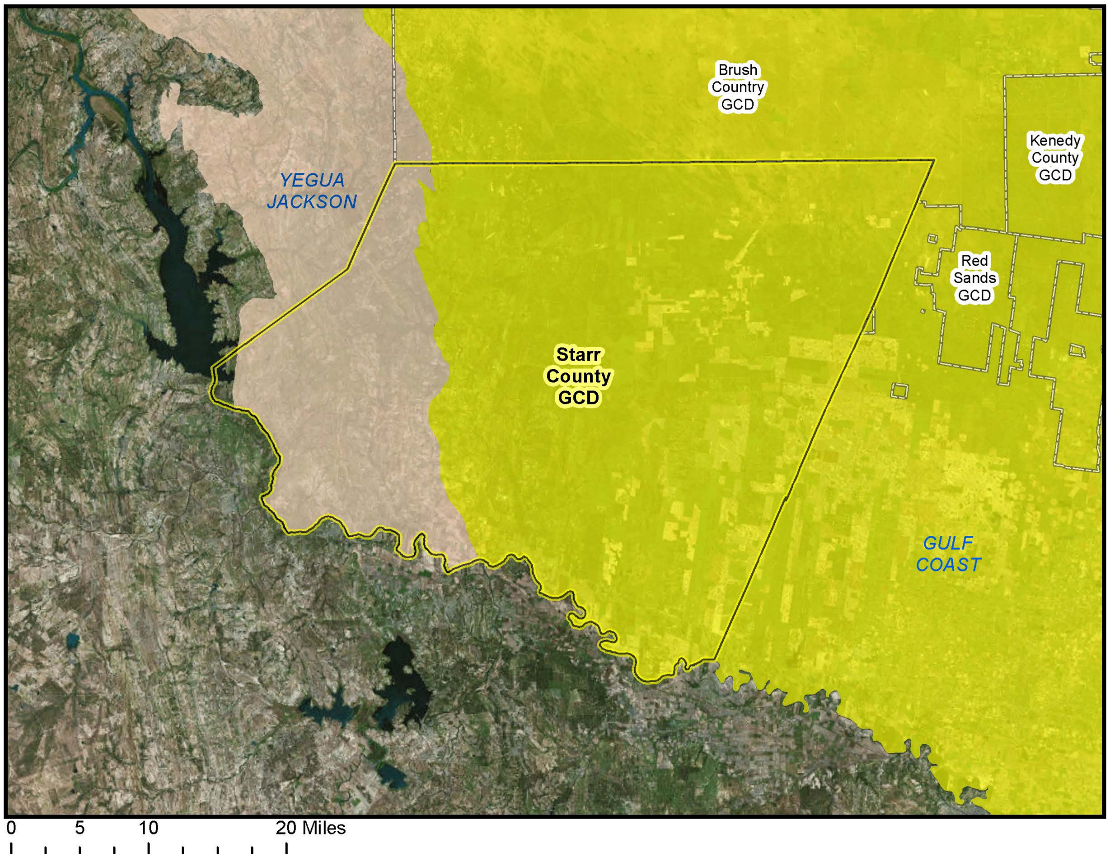
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



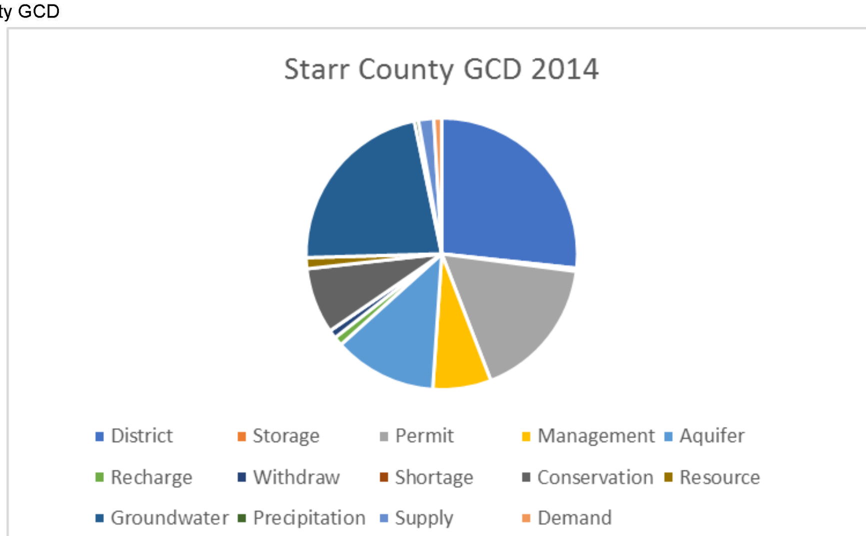
Southwestern Travis County GCD

Southwestern Travis County GCD was created in 2017 and will release its first GCD management plan in the near future.

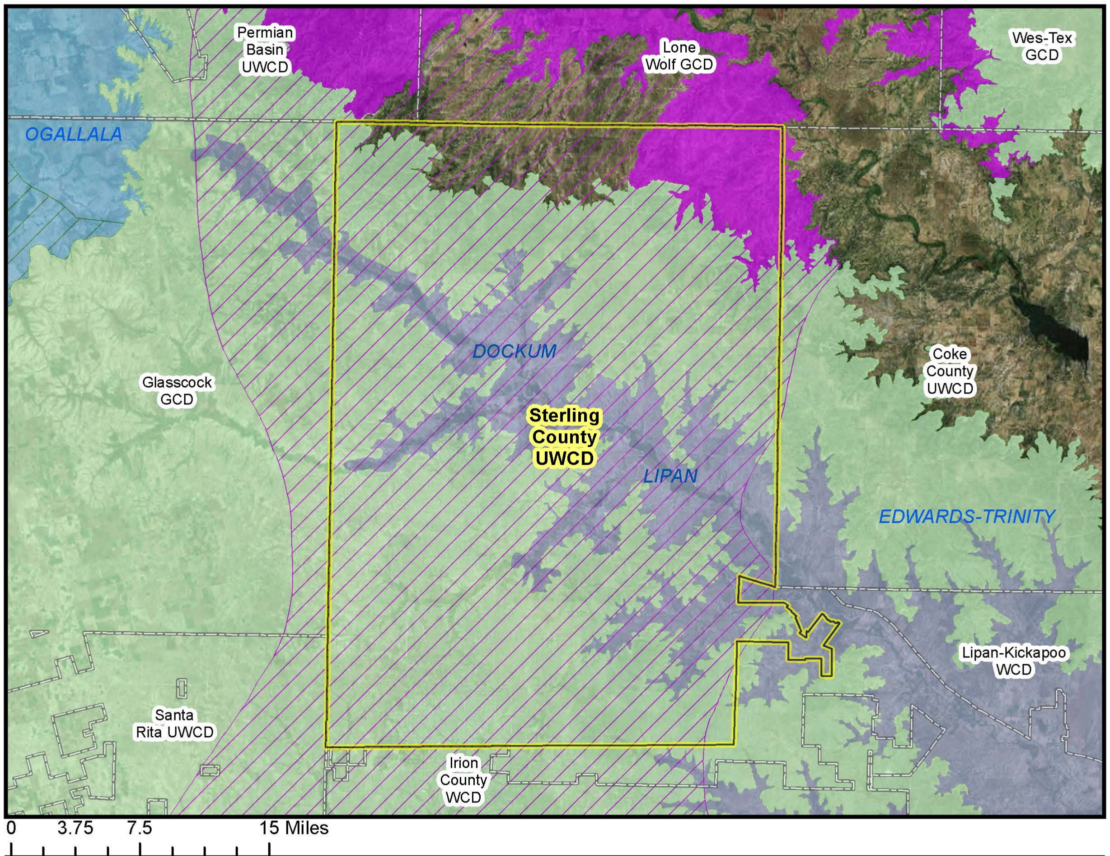
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



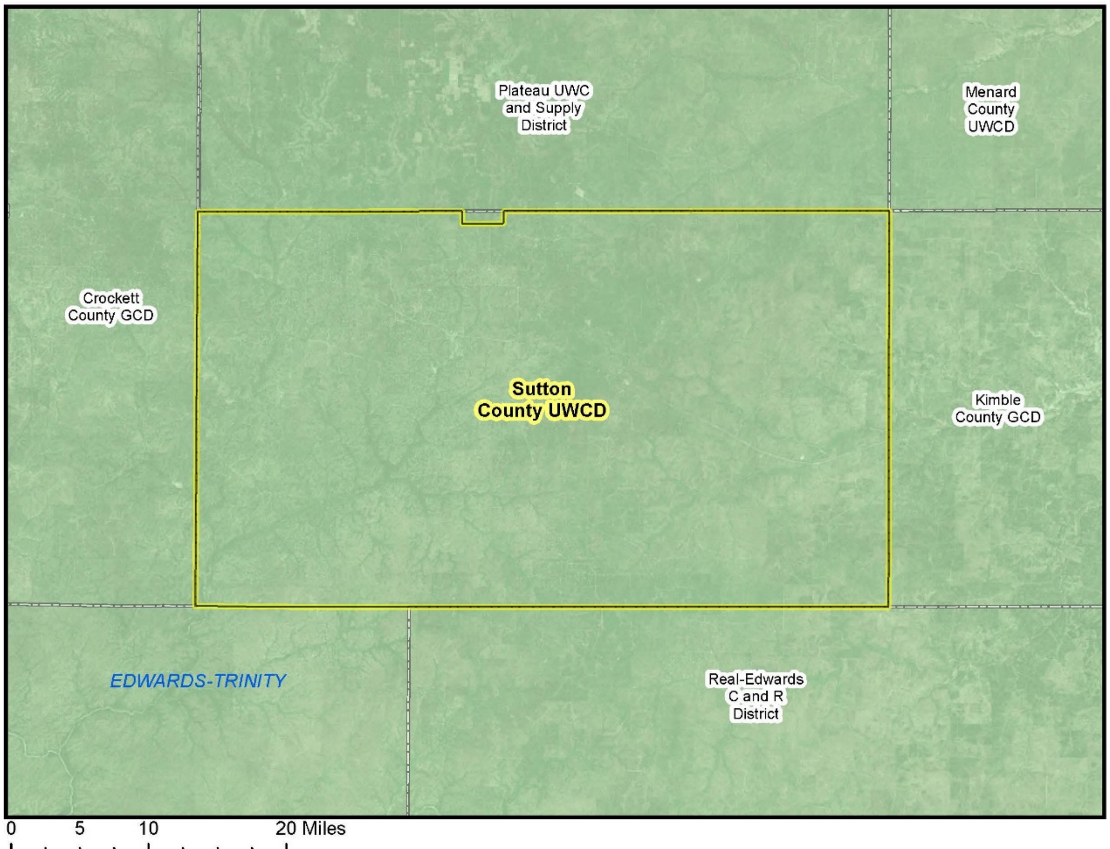
Starr County GCD

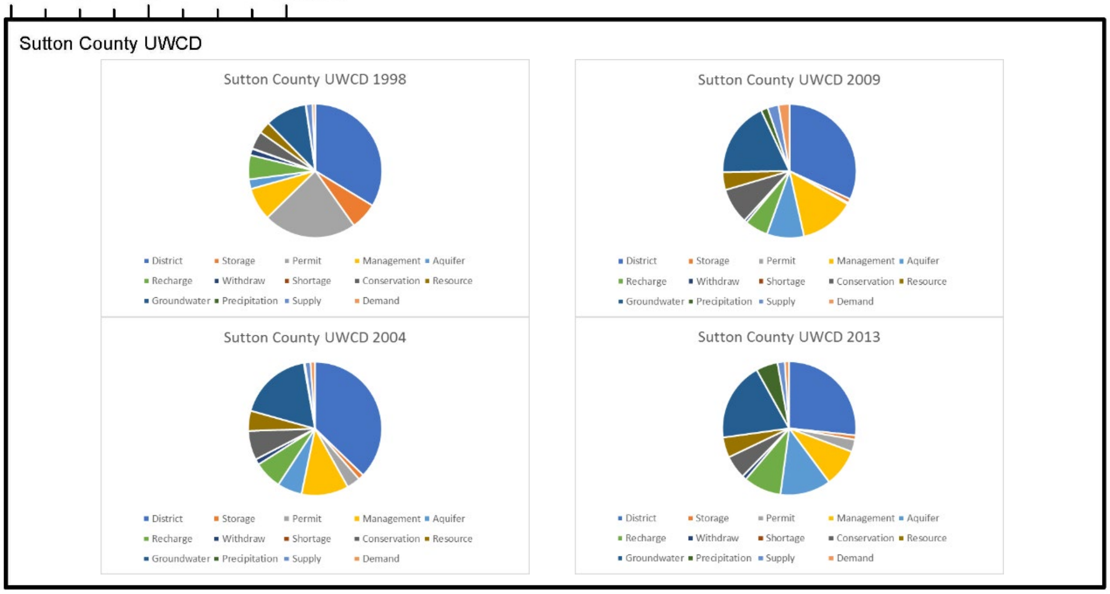


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

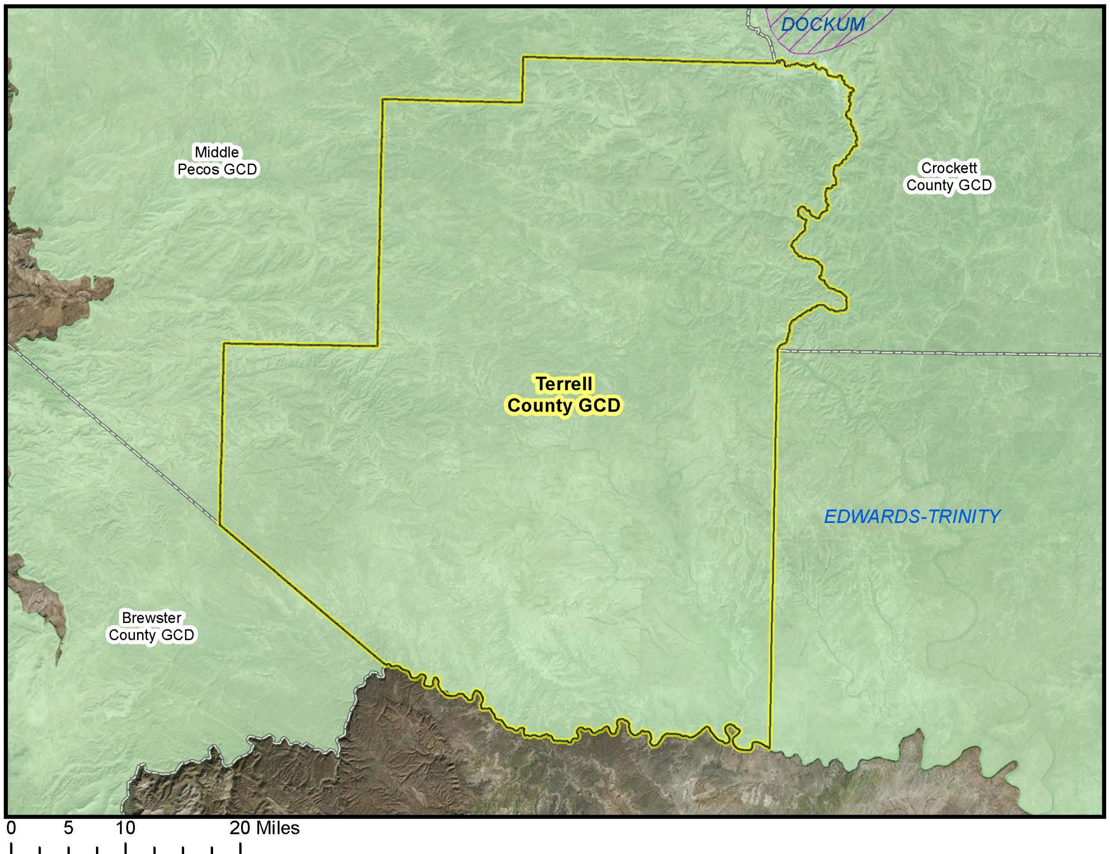


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

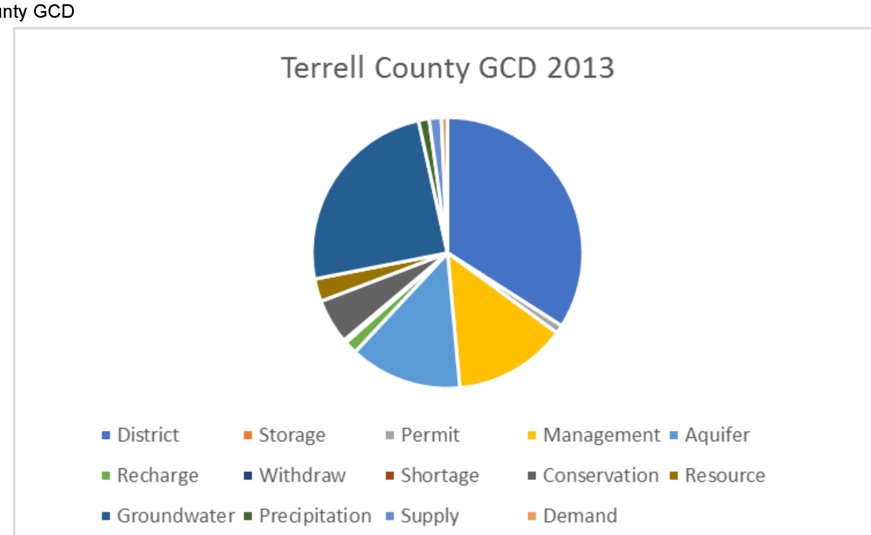




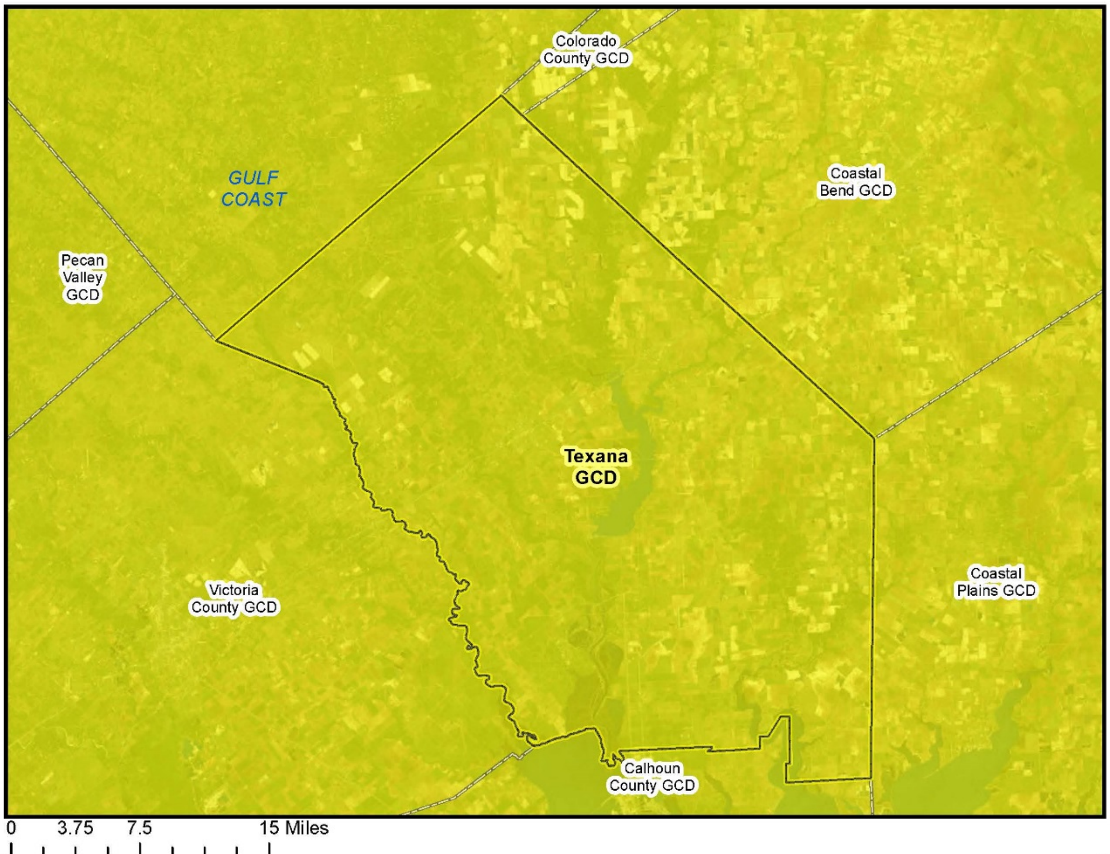
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

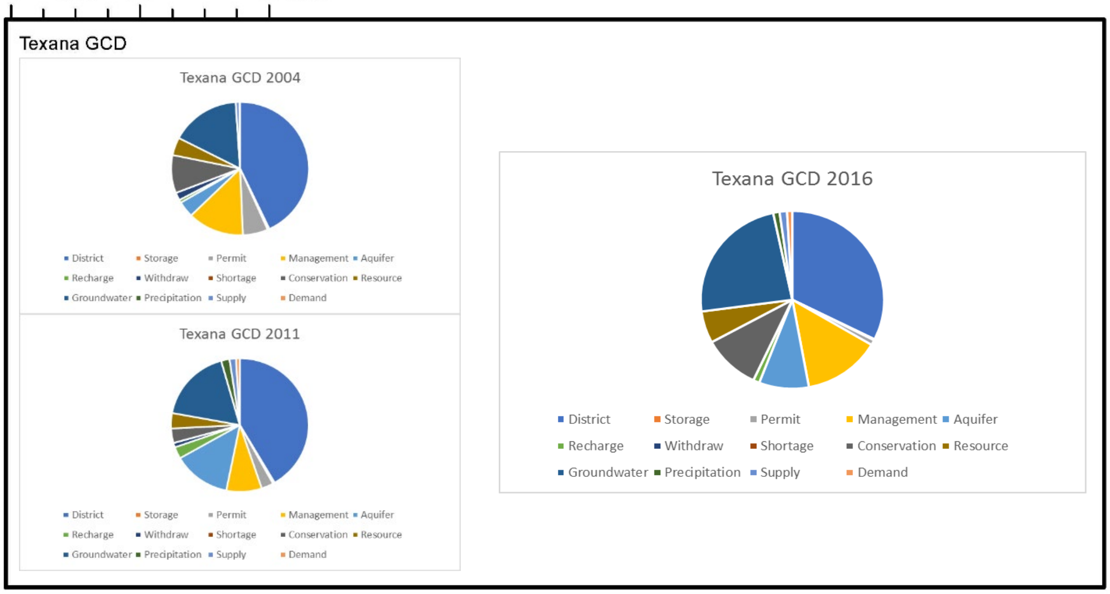


Terrell County GCD

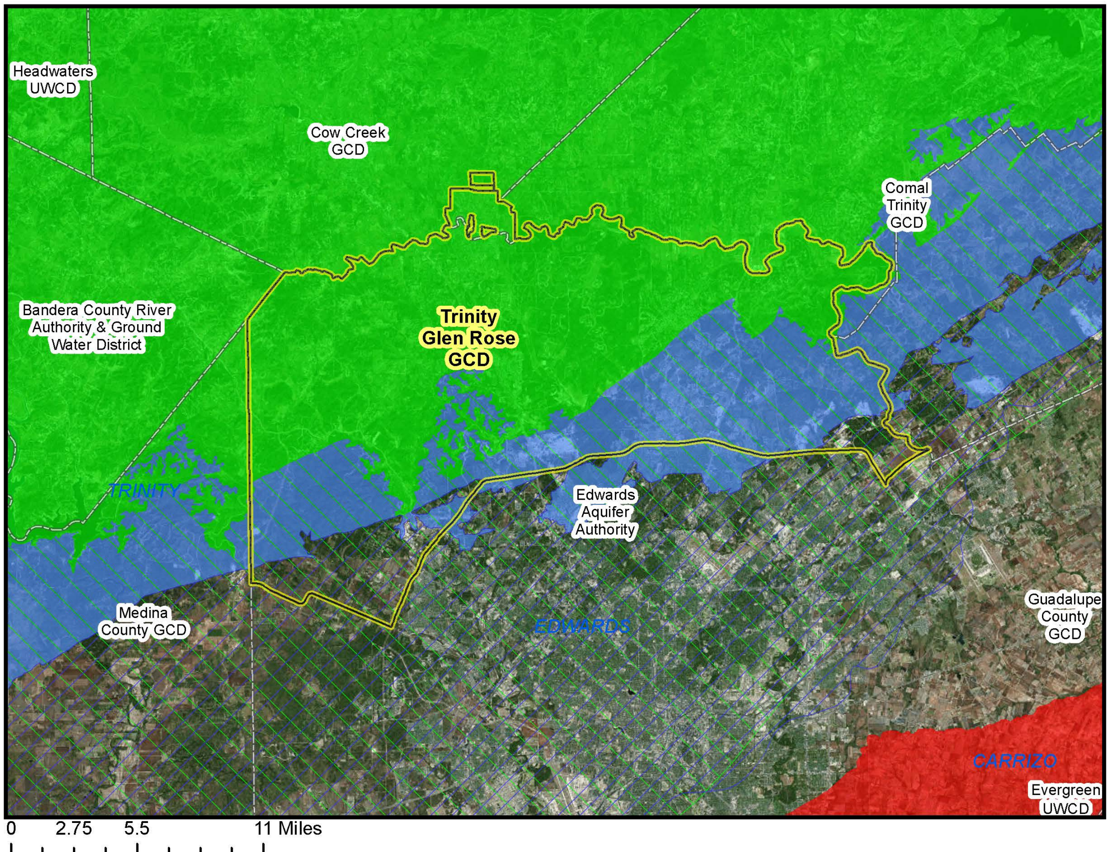


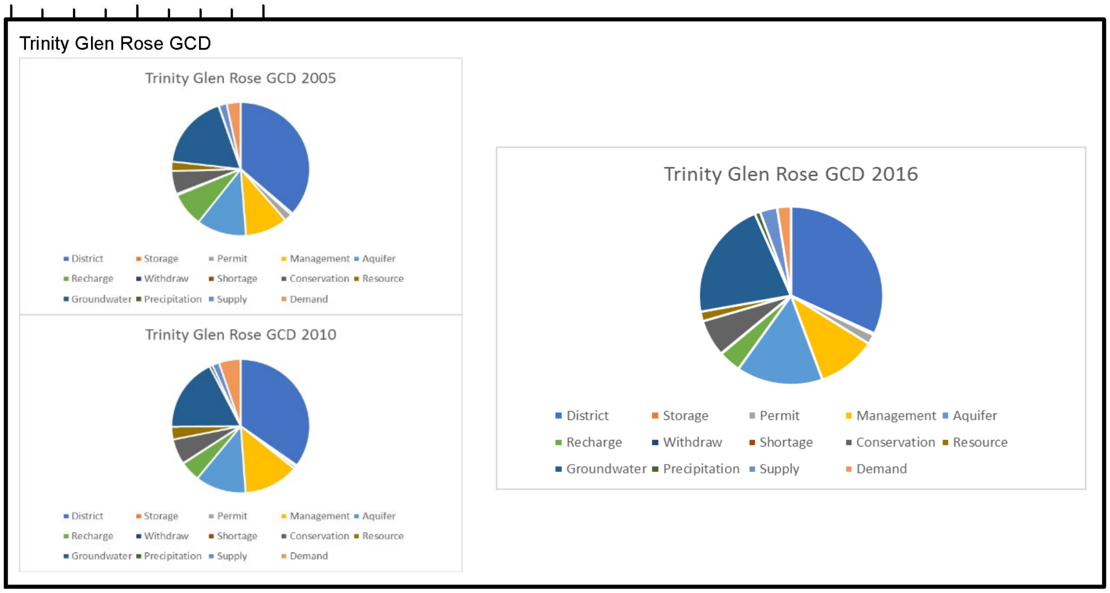
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



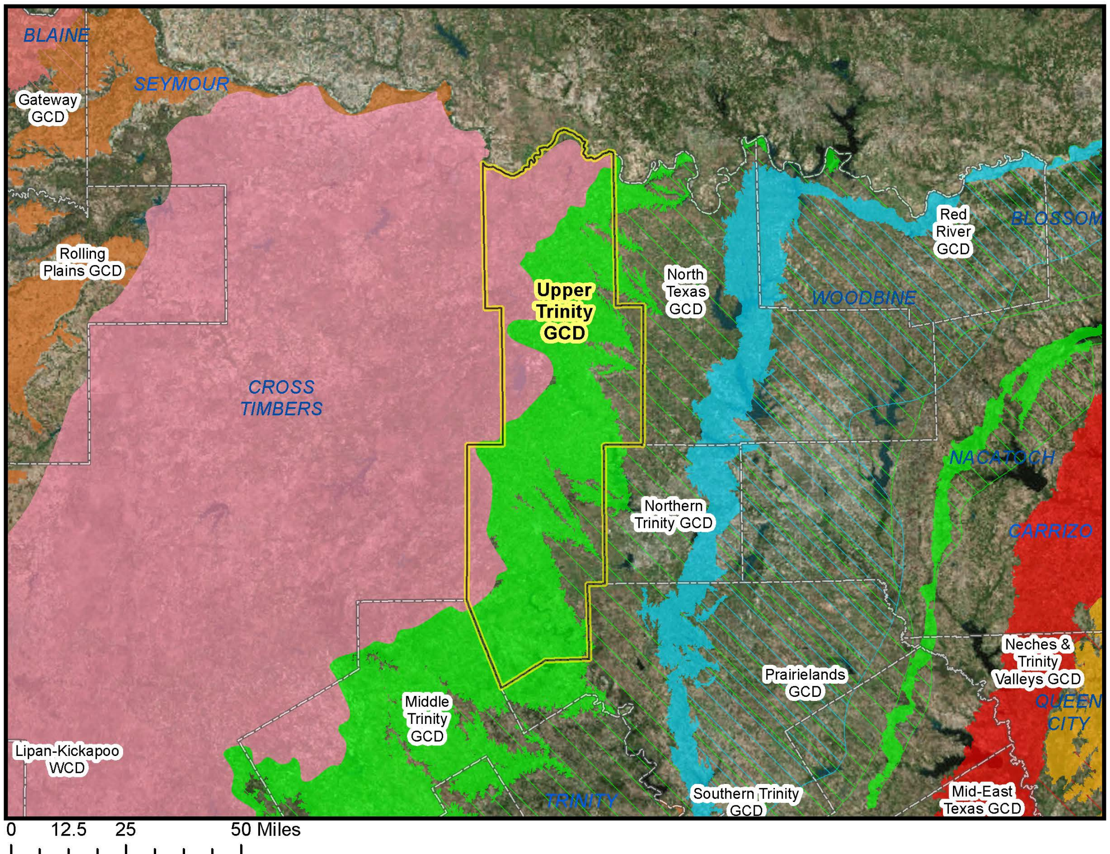


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

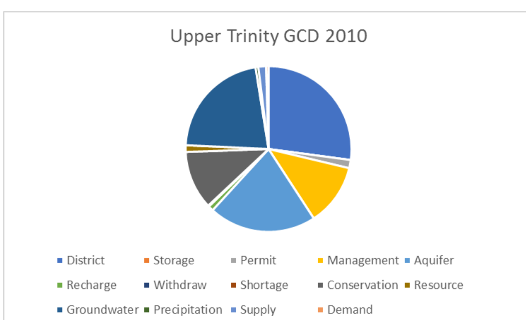


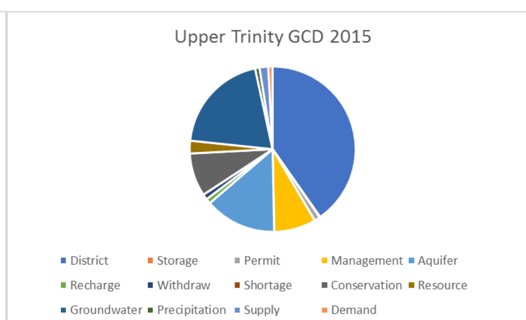


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

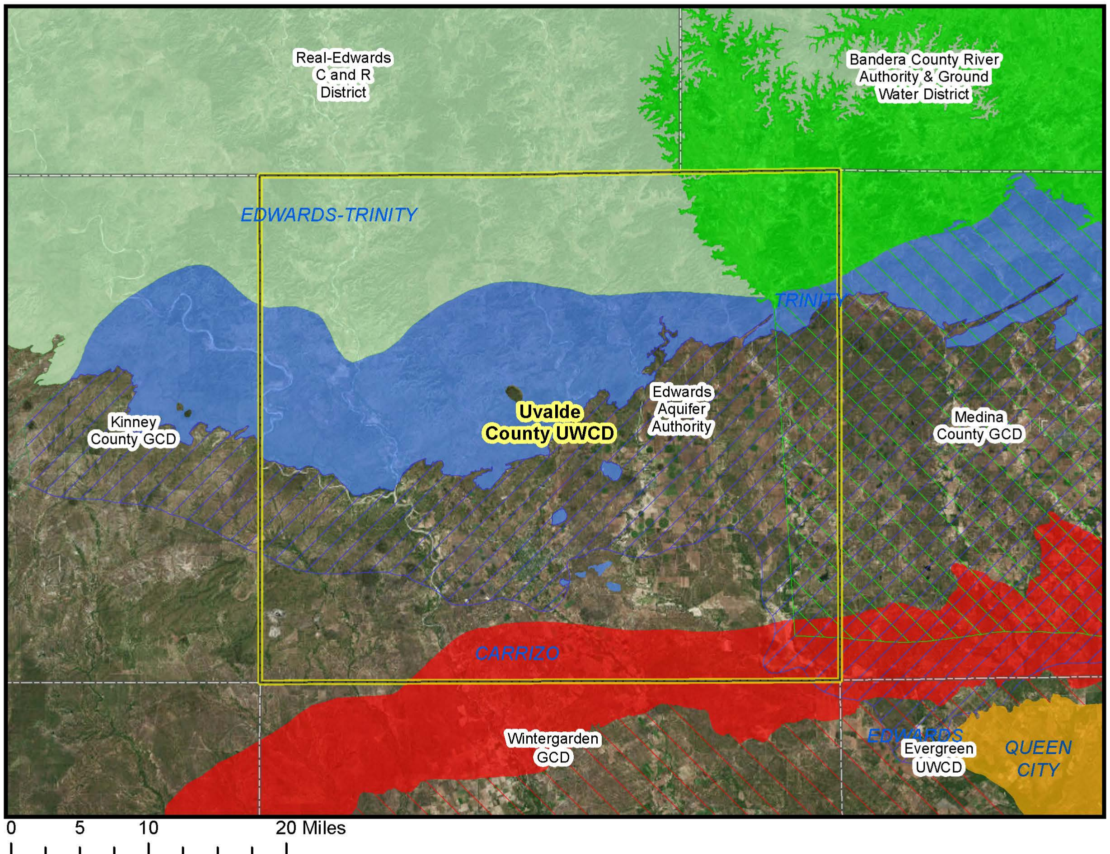


Upper Trinity GCD

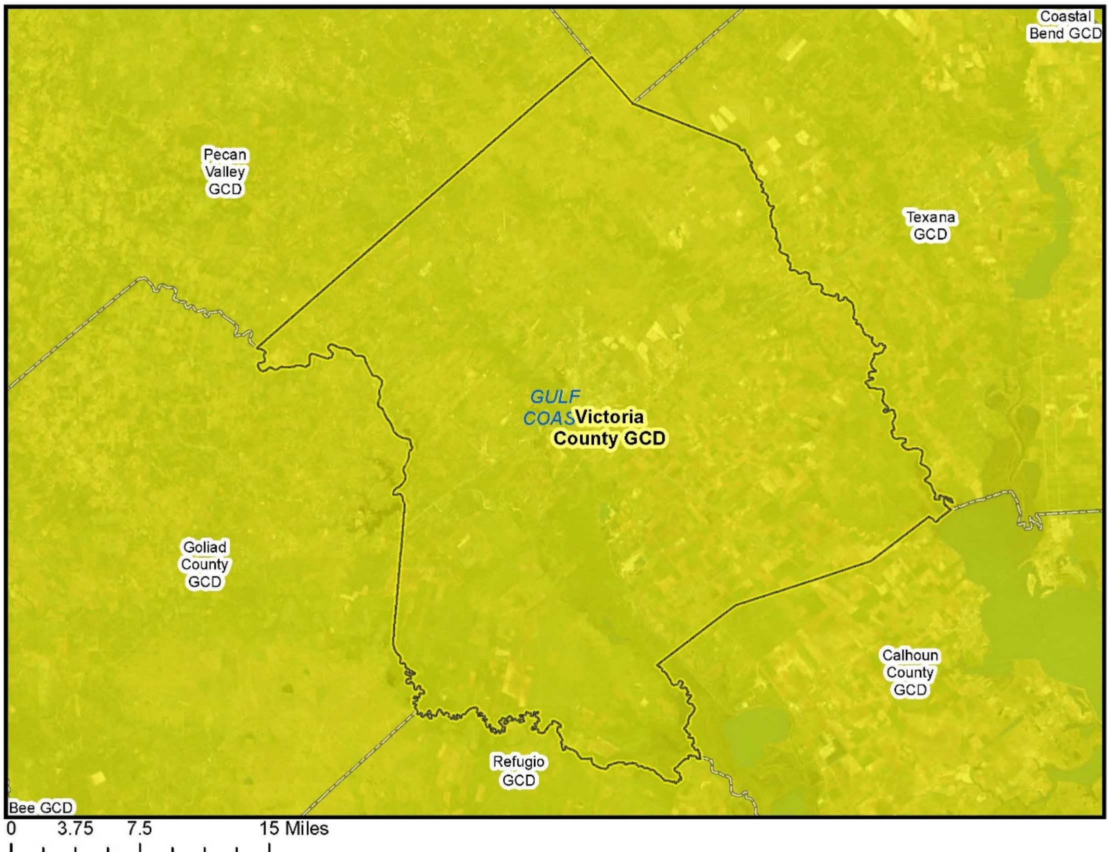


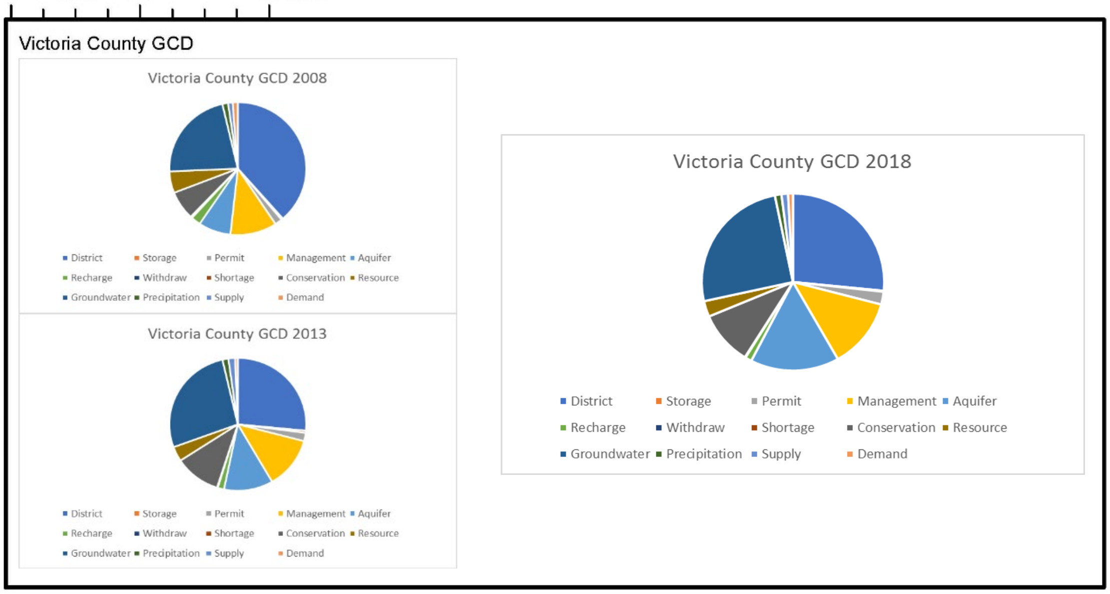


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

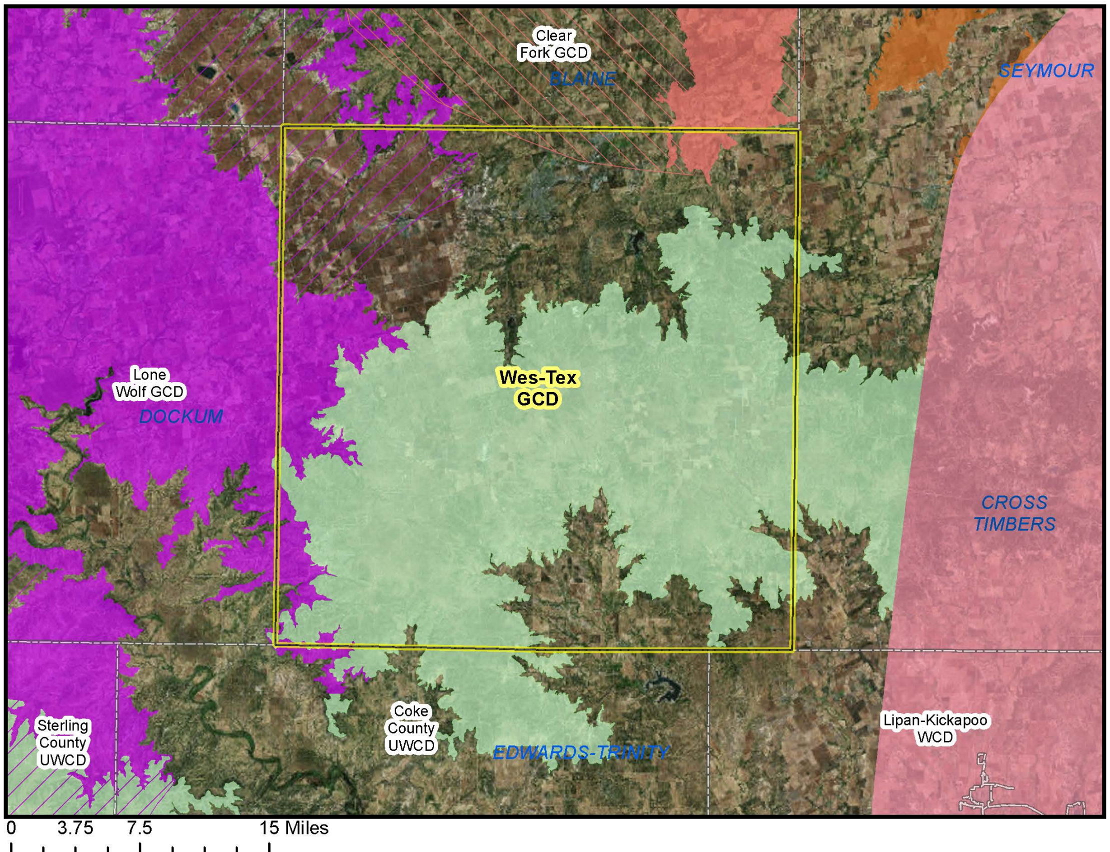


Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

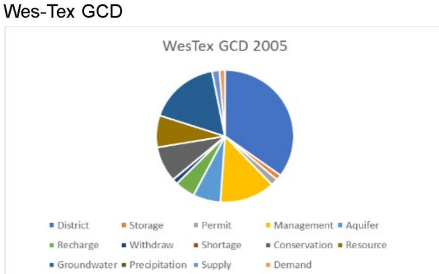


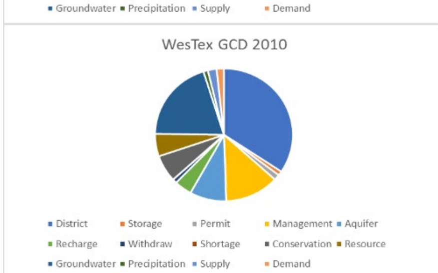


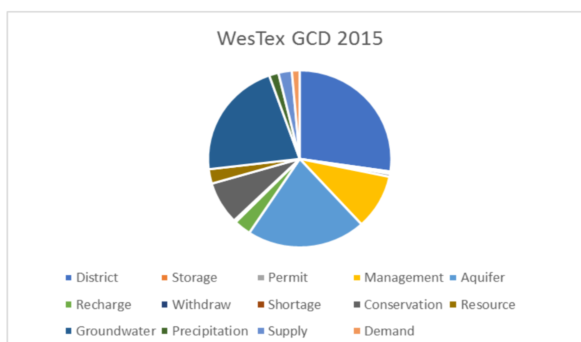
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



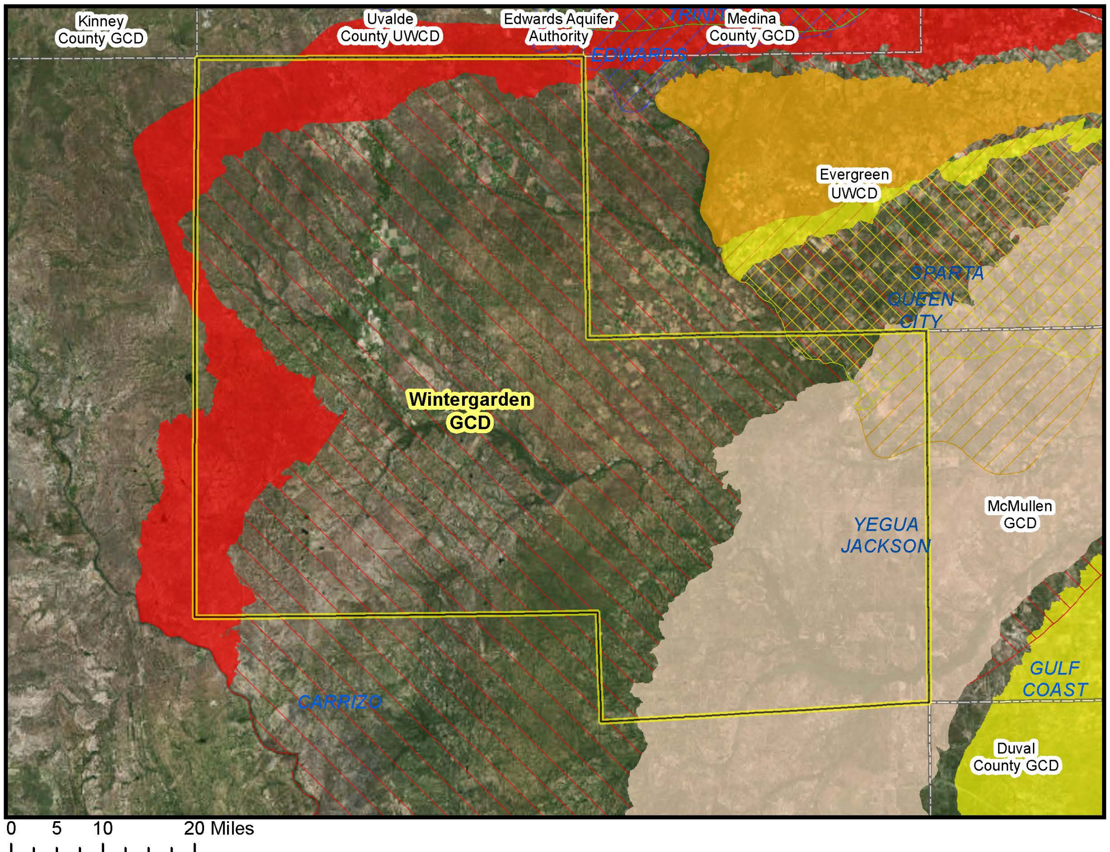
Wes-Tex GCD



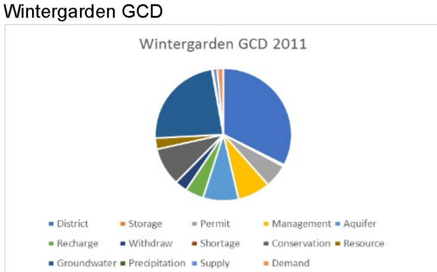


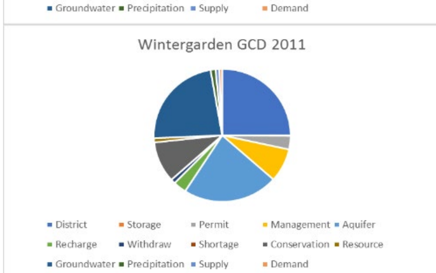


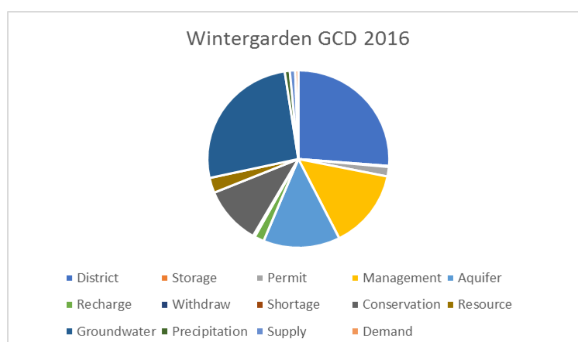
Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community



Wintergarden GCD







Service Layer Credits: Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

List of Acronyms

CERCLA - Comprehensive Environmental Response, Compensation, and Liability Act

DFC – Desired Future Conditions

EU – European Union

FAO - The Food and Agriculture Organization of the United Nations

GAM – Groundwater Availability Model

GCD – Groundwater Conservation District

GMA – Groundwater Management Area

GW-MATE - Groundwater Management Advisory Team

GWP – Global Water Partnership

IAH - International Association of Hydrogeologists

IUCN - International Union for Conservation of Nature

IWRM – Integrated Water Resource Management

IWMI - International Water Management Institute

LDA – Linear Discriminant Analysis

MAG – Modeled Available Groundwater

NGWA - National Ground Water Association

PGMA – Priority Groundwater Management Area

SVM – Support Vector Machine

TCEQ – Texas Commission on Environmental Quality

TWDB – Texas Water Development Board

UNESCO - The United Nations Educational, Scientific and Cultural Organization

UWCD – Underwater Conservation District

UTFI - Underground Taming of Floods for Irrigation

Works Cited

Ayres, Andrew, et al. "How Transaction Costs Obstruct Collective Action: Evidence from California's Groundwater." *Journal of Environmental Economics and Management*, vol. 91, 20 July 2018, pp. 46–65., doi:10.3386/w23382.

Bithas, Kostos and Chrysostomos Stoforos, 2006. "Estimating Urban Residential Water Demand Determinants and Forecasting Water Demand for Athens Metropolitan Area, 2000-2010," South-Eastern Europe Journal of Economics, Association of Economic Universities of South and Eastern Europe and the Black Sea Region, vol. 4(1), pages 47-59.

Bruun, Bech, et al. "Water for Texas 2017 State Water Plan." *Texas Water Development Board*.

Edalat, Farideh Delavari, and M. Reza Abdi. "Adaptive Water Management; Concepts, Principles and Applications for Sustainable Development." *International Series in Operations Research & Management Science*, vol. 258, 2018, doi:10.1007/978-3-319-64143-0.

Edwards, Eric C. "What Lies Beneath? Aquifer Heterogeneity and the Economics of Groundwater Management." Journal of the Association of Environmental and Resource Economists, vol. 3, no. 2, 2016, pp. 453–491., doi:10.1086/685389.

FAO. *Global Framework for Action to Achieve the Vision on Groundwater Governance*. FAO, 2016.

Gun, Jac Van Der, and Jean Margat. "Groundwater around the World." 19 Mar. 2013, doi:10.1201/b13977.

IGRAC. "GW-MATE." International Groundwater Resources Assessment Centre, Water Anchor of the World Bank, www.un-igrac.org/special-project/gw-mate.

James, Gareth, et al. *An Introduction to Statistical Learning with Applications in R*. Springer, 2017.

Loftus, Elizabeth F., and Guido Zanni. "Eyewitness Testimony: The Influence of the Wording of a Question." *Bulletin of the Psychonomic Society*, vol. 5, no. 1, 1975, pp. 86–88., doi:10.3758/bf03336715.

Madani, Kaveh, and Ariel Dinar. "Non-Cooperative Institutions for Sustainable Common Pool Resource Management: Application to Groundwater." *Ecological Economics*, vol. 74, 2012, pp. 34–45., doi:10.1016/j.ecolecon.2011.12.006.

Megdal, Sharon & Zuniga Teran, Adriana & Varady, Robert & Delano, Nathaniel & Gerlak, Andrea & T Vimont, Ethan. (2018). Groundwater governance in the United States: a mosaic of approaches AM-Adaptive Management AMAs-Active Management Areas CAP-Central Arizona Project DWR-Department of Water Resources GMA-Groundwater Management Act IWRM-Integrated Water Resources Management SGMA-Sustainable Groundwater Management Act.

Mullican, William F., and Suzanne Schwartz. "100 Years of Rule of Capture: From East to Groundwater Management." *Report 361 to the Texas Water Development Board*, June 2004, www.twdb.texas.gov/publications/reports/numbered_reports/doc/R361/R361.pdf.

NGWA, "Facts About Global Groundwater Usage." National Ground Water Association, www.ngwa.org/what-is-groundwater/About-groundwater/facts-about-global-groundwater-usage. 2019

Neffendorf, Blake, and Janie Hopkins. "Summary of Groundwater Conditions in Texas: Recent (2014–2015) and Historical WaterLevel Changes in the TWDB Recorder Network ." *Texas Water Development Board, Technical Note 15-07*, Dec. 2015.

Saha, Dipankar, et al. "Clean and Sustainable Groundwater in India." *Springer Hydrogeology*, 2018, doi:10.1007/978-981-10-4552-3.

Smith, Steven M., et al. "Responding to a Groundwater Crisis: The Effects of Self-Imposed Economic Incentives." *Journal of the Association of Environmental and Resource Economists*, vol. 4, no. 4, 2017, pp. 985–1023., doi:10.1086/692610.

Texas Const. art. XVI,§59

Texas Supreme Court. H. T.C. Ry. Co. v. East. 13 June 1904.

81 S.W. 279, 98 Tex. 146

Texas Supreme Court. Friendswood Devel. Co. v. Smith-Southwest Indus., Inc.29 Nov. 1978. 576 S.W.2d 21 (1978)

Title 31, Part 10, §356.10 (6) of the Texas Administrative Code

Wagner, B.J. and Gannett, M.W., 2014, Evaluation of alternative groundwater-management strategies for the Bureau of Reclamation Klamath Project, Oregon and California: U.S. Geological Survey Scientific Investigations Report 2014–5054, 48 p., http://dx.doi.org/10.3133/sir20145054.

Villholth, Karen G., et al. *Advances in Groundwater Governance*. CRC Press/Balkema, 2018.